CREDIT CARD ACCEPTANCE

**SATYAM KUMAR, NIT ANDHRA PRADESH**

Project Objective

Credit evaluation and approval is the process a business or an individual must go through to become eligible for a loan or to pay for goods and services over an extended period. It also refers to the process businesses undertake while evaluating a request for credit.

Granting credit approval depends on the willingness of the creditor to lend money in the current economy and that same lender's assessment of the ability and willingness of the borrower to return the money or pay for the goods obtained plus interest in a timely fashion.

Typically, small businesses must seek credit approval to obtain funds from lenders, investors, and vendors, and also grant credit approval to their customers.

The project involves the use of various machine learning algorithms and identify best algorithm for better performance. This improves the accuracy of credit card approval.

Project Scope

The broad scope of the Credit Card Acceptance project includes:

* Pre-processing the dataset.
* Drawing the various graphs like countplot ,barplot,scatter plot.
* To split dataset into training and testing in proper threshold limit.
* Feature Extraction Selection of classifier.
* Random Forest Classifier, Decision Tree Classifier,

K-Nearest Neighbours Classifier has been used in proposed model .

Data description

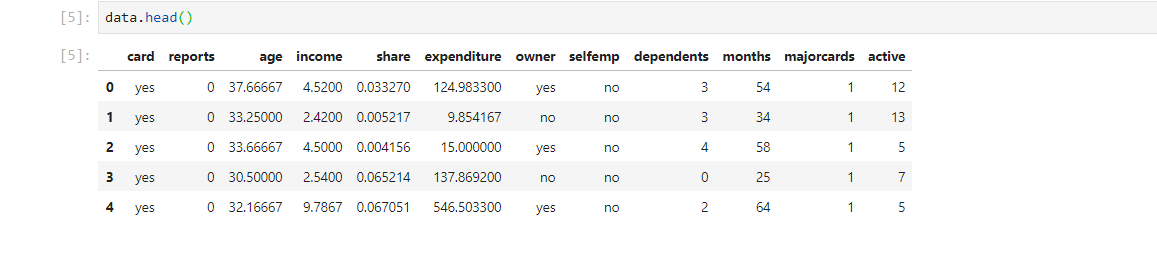
• The small credit card dataset is one of datasets for simple econometric analysis taken from Kaggle, originally from William Greene's book Econometric Analysis.

The code below will load the credit card dataset.

1. data=pd.read\_csv("D:\creditcard.csv")
2. data.head()
3. print(data.columns)

4.print(data.shape)

Outputs:

* 

Index(['card', 'reports', 'age', 'income', 'share', 'expenditure', 'owner',

'selfemp', 'dependents', 'months', 'majorcards', 'active'],

dtype='object')

* (1319,12)

Model Building

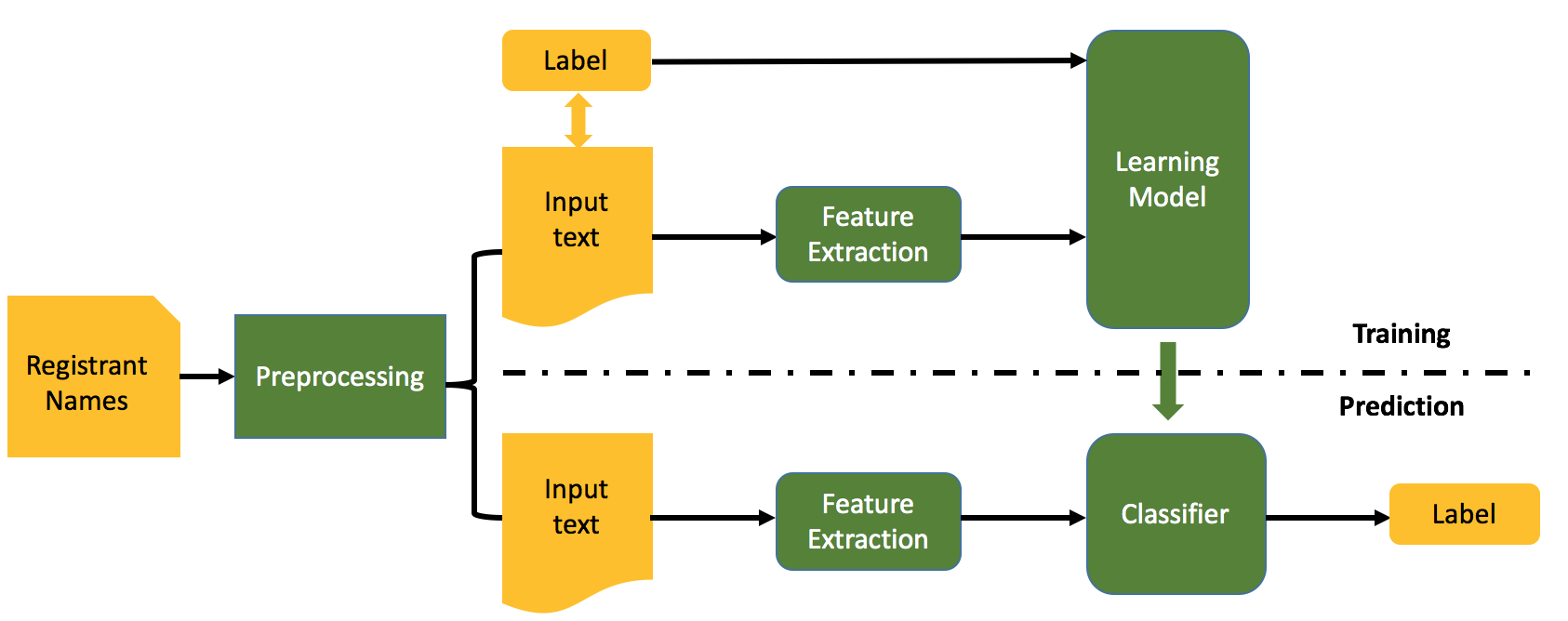
* decision tree
* k-nearest neighbors
* naive bayes
* LOGISTIC regression

**DECISION TREE MODEL**

* A decision tree is a supervised predictive model. It can learn to predict discrete or continuous outputs by answering questions based on the values of the inputs it receives.
* It is a graphical representation of all the possible solutions to a decision based on certain conditions.
* It is based on CART algorithm(Classification And Regression Tree Algorithm)
* Highly interpretable.
* Can be specified as a series of rules.
* More closely approximates human decision-making than other ML algorithms.
* Fast prediction.
* Features don’t need scaling.

## 

## Work Flow



## Features Selection

The selected feature attributes and their properties:

* Feature attributes and their scores:

[('reports', 0.004721763304312109),

('expenditure', 0.9783798350814344),

('age', 0.010235979525498706),

('income', 0.0066624220887548545)]

K-NN(K-Nearest Neighbours)

• A powerful classification algorithm used in pattern recognition.

• K nearest neighbors stores all available cases and classifies new cases based on a similarity measure( e. g. distance function ).

• One of the top data mining algorithms used today.

• A non-parametric lazy learning algorithm (An instance based Learning method).

KNN: Classification Approach

• An object (a new instance) is classified by a majority votes for its neighbor classes.

• The object is assigned to the most common class amongst its K nearest neighbors.(measured by a distant function )

Strength of KNN

• Very simple and intuitive.

• Can be applied to the data from any distribution.

• Good classification if the number of samples is large enough.

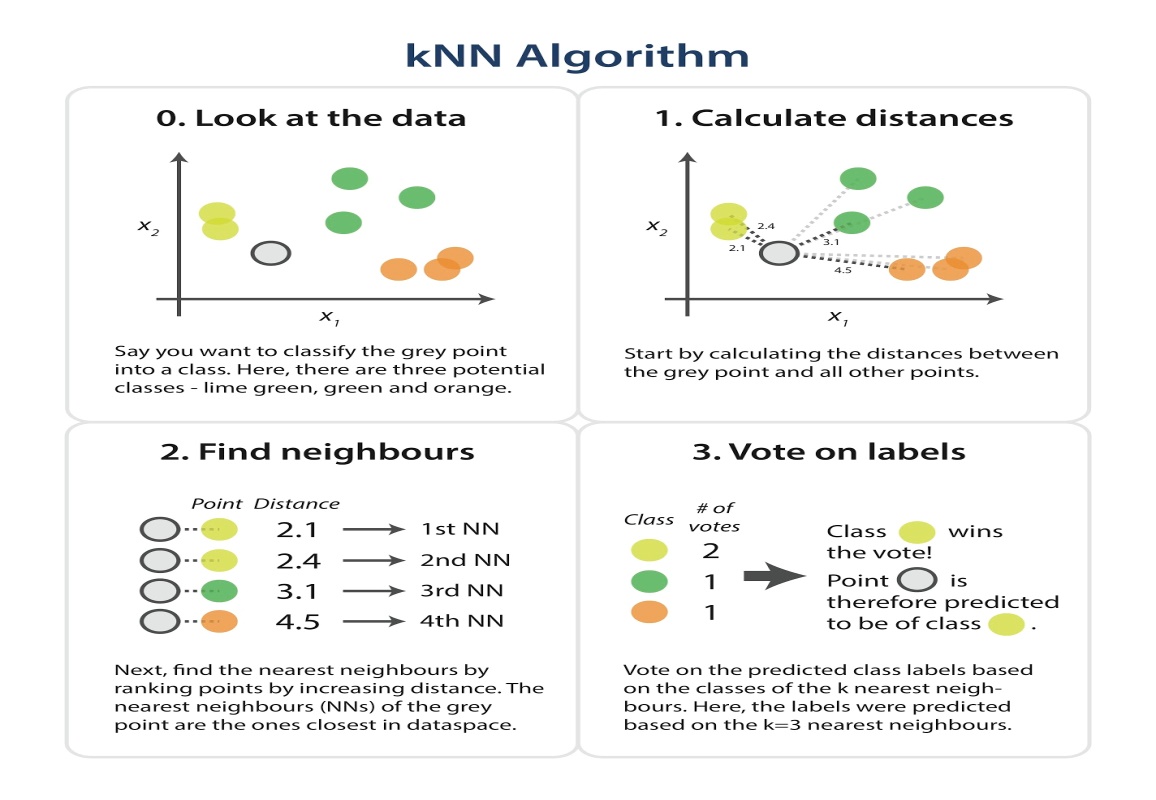
Weakness of KNN

• Takes more time to classify a new example.

• Need to calculate and compare distance from new example to all other examples.

• Choosing k may be tricky.

• Need large number of samples for accuracy.



NAIVE BAYES MODEL

INTRODUCTION

Naive Bayes methods are a set of supervised learning algorithms based on applying Bayes’ theorem with the “naive” assumption of conditional independence between every pair of features given the value of the class variable. Bayes’ theorem states the following relationship, given class variable y and dependent feature vector x1 through xn, :



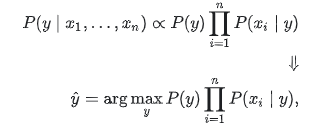
Using the naive conditional independence assumption that



for all i, this relationship is simplified to



Since P(x1,…,xn) is constant given the input, we can use the following classification rule:



and we can use Maximum A Posteriori (MAP) estimation to estimate P(y) and P(xi∣y); the former is then the relative frequency of class y in the training set.

The different naive Bayes classifiers differ mainly by the assumptions they make regarding the distribution of P(xi∣y).

In spite of their apparently over-simplified assumptions, naive Bayes classifiers have worked quite well in many real-world situations, famously document classification and spam filtering. They require a small amount of training data to estimate the necessary parameters.

Although naive Bayes is known as a decent classifier, it is known to be a bad estimator, so the probability outputs from predict\_proba are not to be taken too seriously.

BAYES THEOREM:

* P(class/data)=P(data/class)\*P(class)/P(data)

Where

* + Class is a particular class
  + Data is an observation’s data
  + P(class/data) is called the posterior
  + P(data/class) is called the likelihood
  + P(class) is called the prior
  + P(data) is called the marginal probability

OUTCOME

* Posterior =(prior x likelihood)/evidence

MODEL IMPLEMENTATION

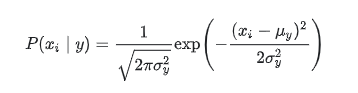
TYPES OF NAIVE BAYES MODEL:

* Gaussian Naïve Bayes
* Multinomial Naïve Bayes
* Bernoulli Naïve Bayes

Here we are going to implement Gaussian Naïve Bayes Model.

GAUSSIAN NAIVE BAYES MODEL

The formula to calculate the likelihood probability is given below. The likelihood of the features is assumed to be Gaussian:



The parameters σy and μy are mean of the data and variance of data respectively.

PREPROCESSING OF DATA

* It is common to find that several attributes in dataset are useless.
* Pre-processing can be done in several steps as importing libraries, importing the dataset, encoding of categorical values, splitting the data-set into Training and Test data-set.

FEATURE SELECTION

Feature selection is one of the most important pre-processing steps in data mining. It is an effective dimensionality reduction technique to remove noise feature. In general, the basic idea of feature selection algorithm to searches through all possible combinations of attributes in the data to find which subset of features works best for prediction.

Thus, the attribute vectors can be reduced in number by which the most meaningful ones are kept and the irrelevant or redundant ones are removed and deleted.

MODEL EVALUATION:

To test and evaluate the model, 70% of the dataset are used. Instances are extracted and then served as a benchmarking dataset for machine learning problems. By comparing the actual class of the instance with the predicted one (i.e. generated by the classification model), system performance can be measures in term of recall, precision, and F-measure.

These can be mathematically defined as below.



ADVANTAGES:

* Easy to implement
* Requires a small amount of training data to estimate the parameters.
* Good results obtained in most of the cases

DISADVANTAGES

* Assumption: class conditional independence, therefore loss of accuracy
* Practically, dependencies exist among variables

E.g.,Symptoms: fever,cough etc., Disease: lung cancer, diabetes,etc

TYPICAL APPLICATION

* Credit approval
* Face recognition
* Sentiment analysis
* Email spam detection
* Target marketing
* Medical diagnosis

LOGISTIC REGRESSION MODEL:

Generally, in Machine Learning problems can be split into two types:

* Unsupervised Machine Learning
* Supervised Machine Learning

When you are working with unsupervised learning, typically you have dataset of some sort or work and another. What we do is that we apply machine learning and various statistical Methods in order to infer structure about the data, get some deeper understanding about what is going on. We don’t have any Label so we can’t make predictions from it.

In Supervised learning, we have a dataset and some list of possible outcomes and what types of outcome we have, give us regression or the classification problems.

The two main variables in an experiment are the independent and dependent variable.

An [**independent variable**](https://www.thoughtco.com/definition-of-independent-variable-605238) is the variable that is changed or controlled in a scientific experiment to test the effects on [the dependent variable](https://www.thoughtco.com/definition-of-dependent-variable-604998).

A **dependent variable** is the variable being tested and measured [in a scientific experiment](https://www.thoughtco.com/design-science-fair-experiment-606827).

The dependent variable is 'dependent' on the independent variable. As the experimenter changes the [independent variable](https://www.thoughtco.com/understanding-variables-in-science-609060), the effect on the variables observed and recorded.

REGRESSION:

**Regression** is a statistical method used to describe the nature of the relationship between variables, that is, positive or negative, linear or nonlinear.

It is used when the prediction for continuous number has to be done.

The outcome is a positive real number.

Regression Technique vary from Linear Regression to SVR and Random Forest Regression.

* Simple Linear Regression
* Multiple Linear Regression
* Polynomial Regression
* Support Vector for Regression(SVR)

Regression models, both Linear and Non- Linear are used for predicting a real value, like salary for example.

CLASSIFICATION:

When there is a Classification Problem in the context of Machine Learning, typically our outputs are Classes or categories. They are represented by whatever categorical data points we find relevant or just simple integers which represent each individual class.

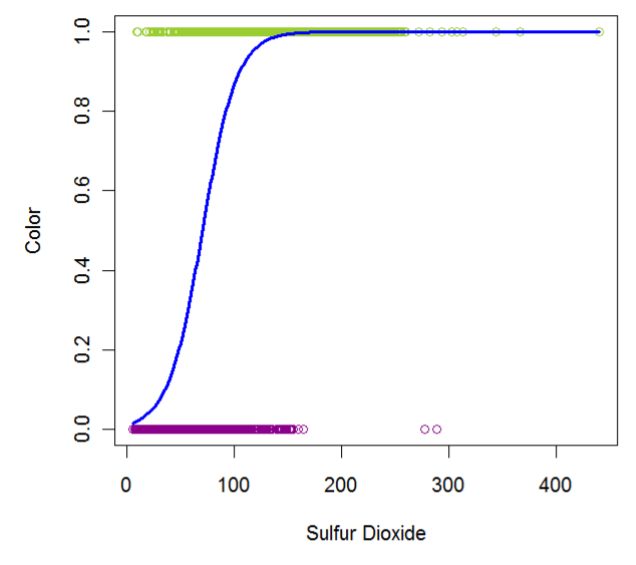
Classification Examples

• Y: presence/absence of disease X: diagnostic measurements

• Y: land cover (grass, trees, water, roads…) X: satellite image data (frequency bands)

• Y: loan defaults (yes/no) X: credit score, own or rent, age, marital status,…

• Y: dementia status X: scores on a battery of psychological tests



Unlike regression where we predict a continuous number, you use classification to predict a category. There is a wide variety of classification applications from medicine to marketing. Classification models include linear models like Logistic Regression, SVM, and nonlinear ones like K-NN, Kernel SVM and Random Forests.

**LOGISTIC REGRESSION**:

Logistic regression analysis studies the association between a categorical dependent variable and a set of independent (explanatory) variables. The name logistic regression is used when the dependent variable has only two values, such as 0 and 1 or Yes and No. The name multinomial logistic regression is usually reserved for the case when the dependent variable has three or more unique values, such as Married, Single, Divorced, or Widowed. Although the type of data used for the dependent variable is different from that of multiple regression, the practical use of the procedure is similar. In the context of low-dimensional data (i.e. when the number of covariates is small compared to the sample size), logistic regression is considered a standard approach for binary classification.

DIMENSIONALITY REDUCTION:

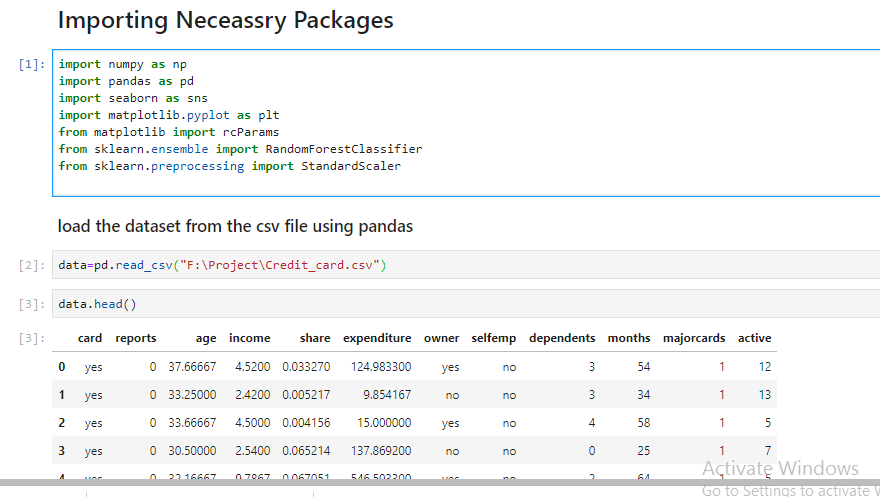
* **Dimensionality reduction** is the process of reducing the number of random variables under consideration by obtaining a set of principal variables. It can be divided into feature selection and feature extraction.
* Feature selection approaches try to find a subset of the original variables (also called features or attributes). It is about choosing some of features based on some statistical score.
* Feature extraction transforms the data in the high-dimensional space to a space of fewer dimensions. It is using techniques to extract some second layer information from the data e.g. interesting frequencies of a signal using Fourier transform.
* Dimensionality reduction helps in data compression, reduces computation time and removes redundant features.

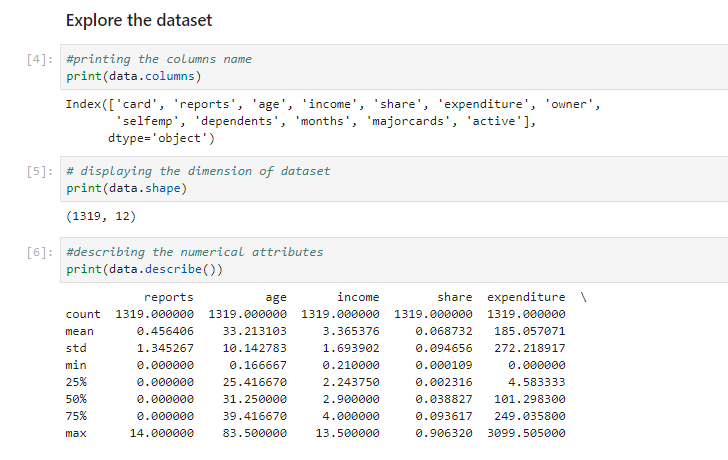
In the given dataset, the feature selection is done using the help of Random Forest Classifier. he random forest (RF) is an “ensemble learning” technique consisting of the aggregation of a large number of decision trees, resulting in a reduction of variance compared to the single decision trees.

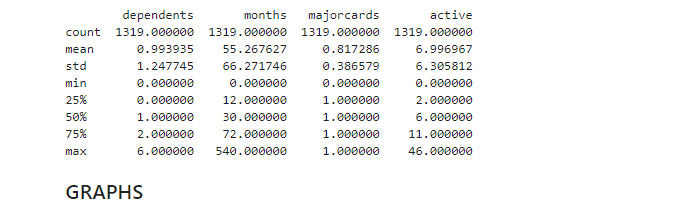
The practical Demonstration of Logistic Regression in our given Dataset for credit card acceptance prediction is illustrated in the Code part of this Documentation.

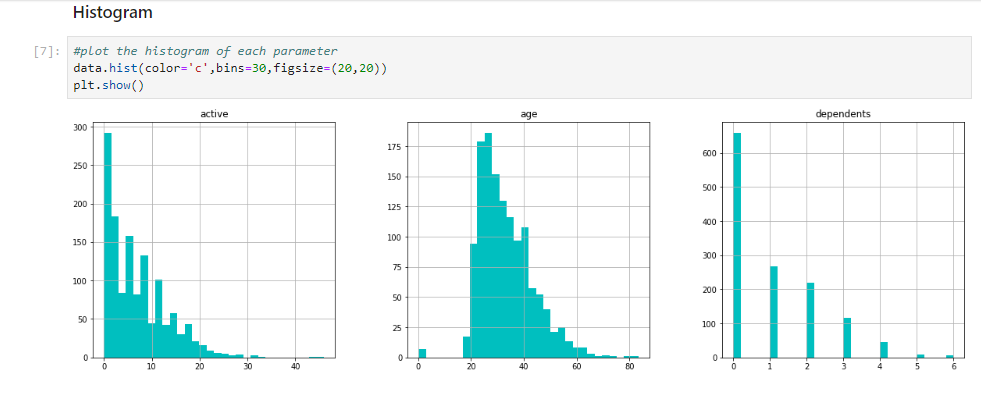
**Code:**

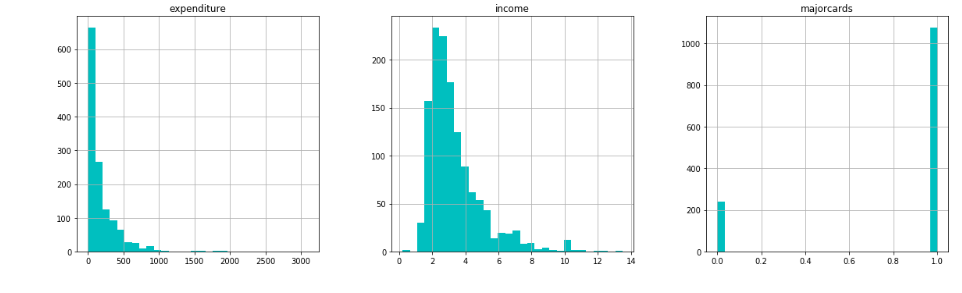
Pre-processing:

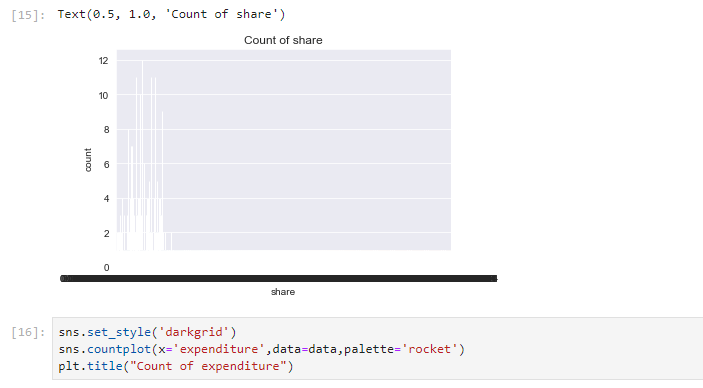
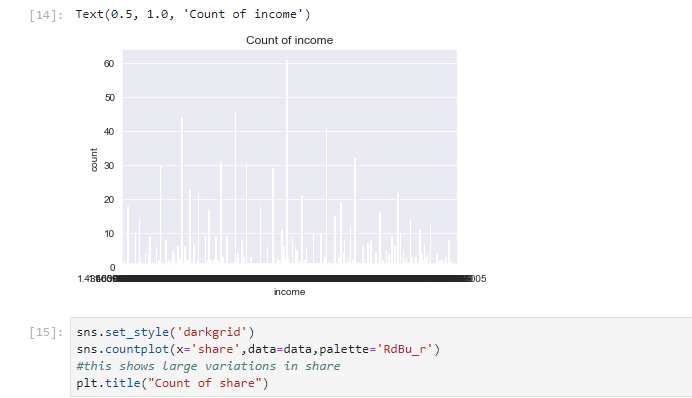
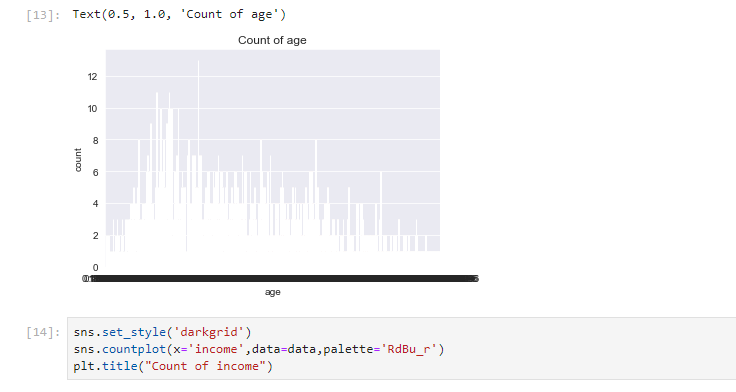
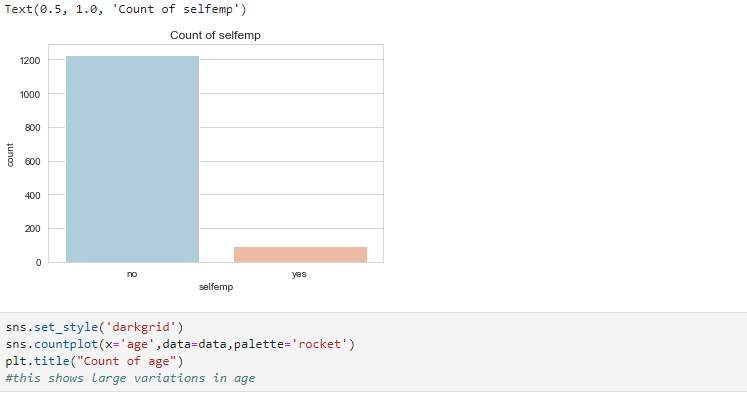
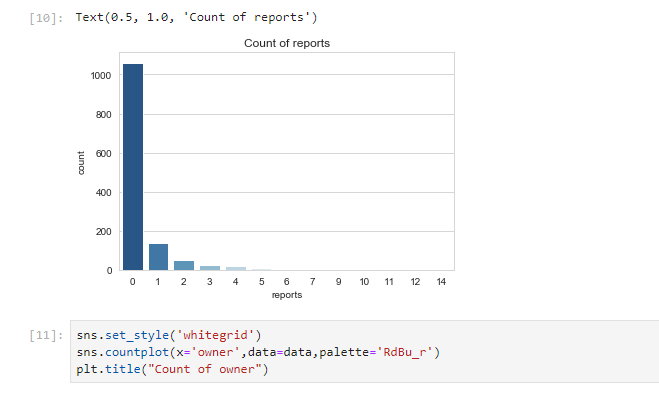
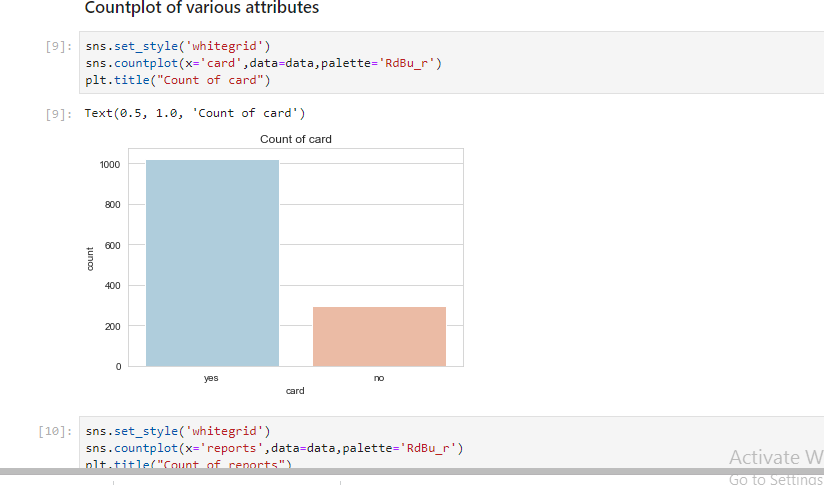
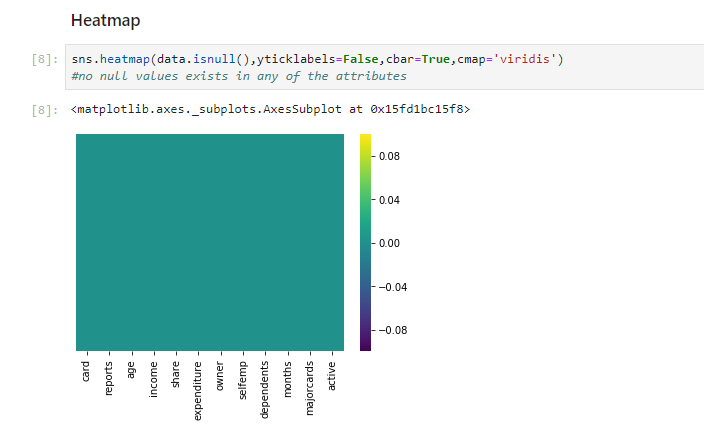
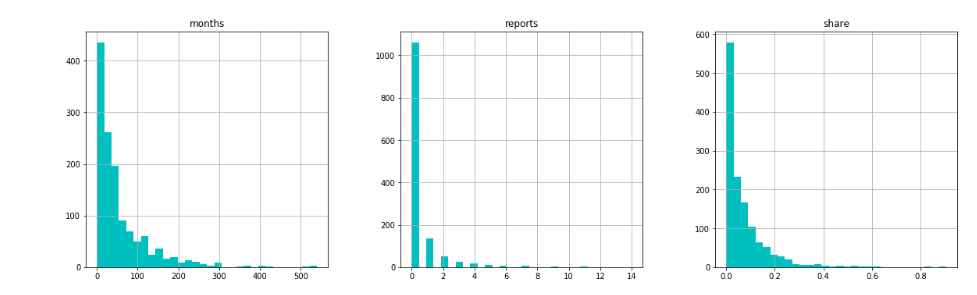
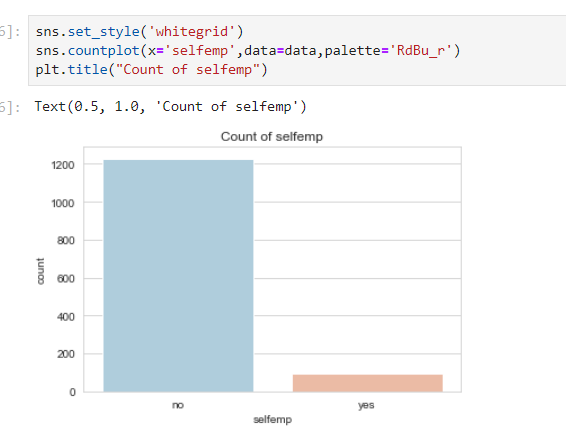
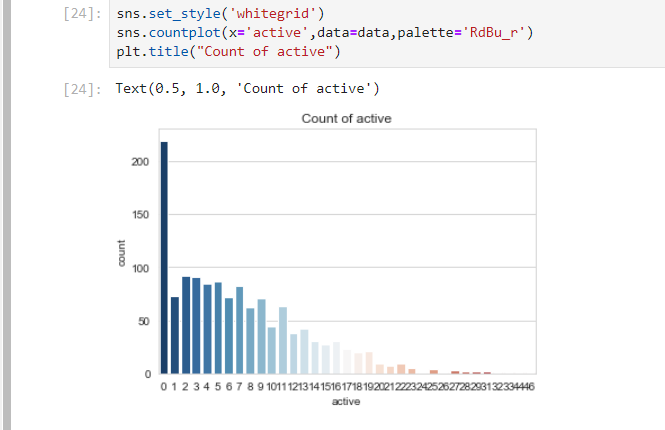
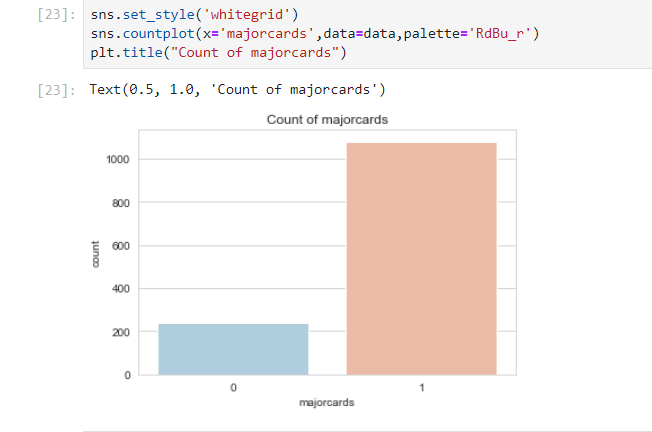
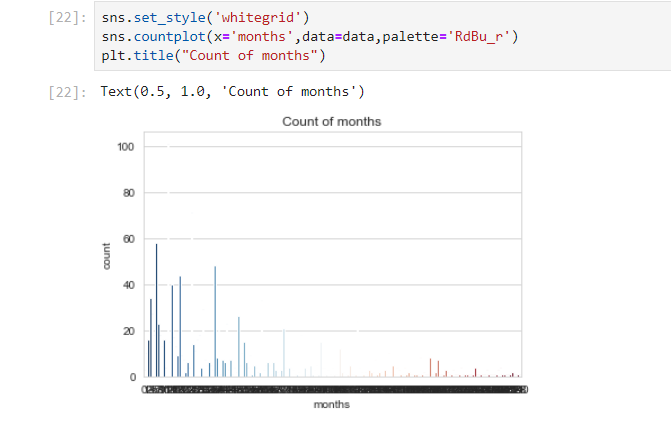
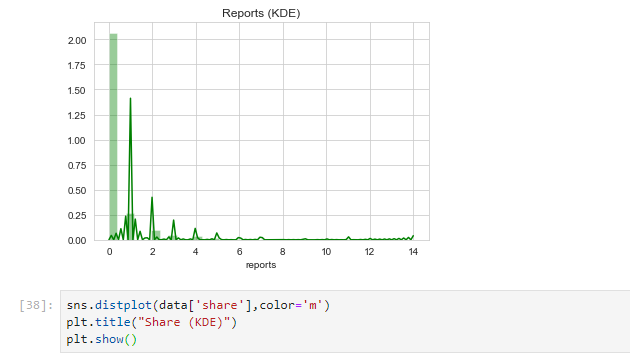
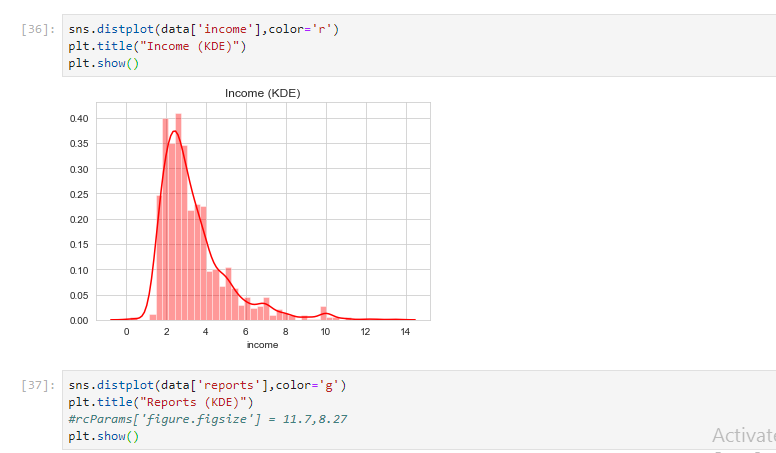
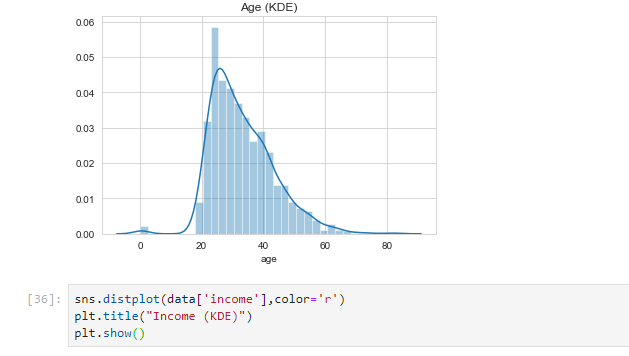
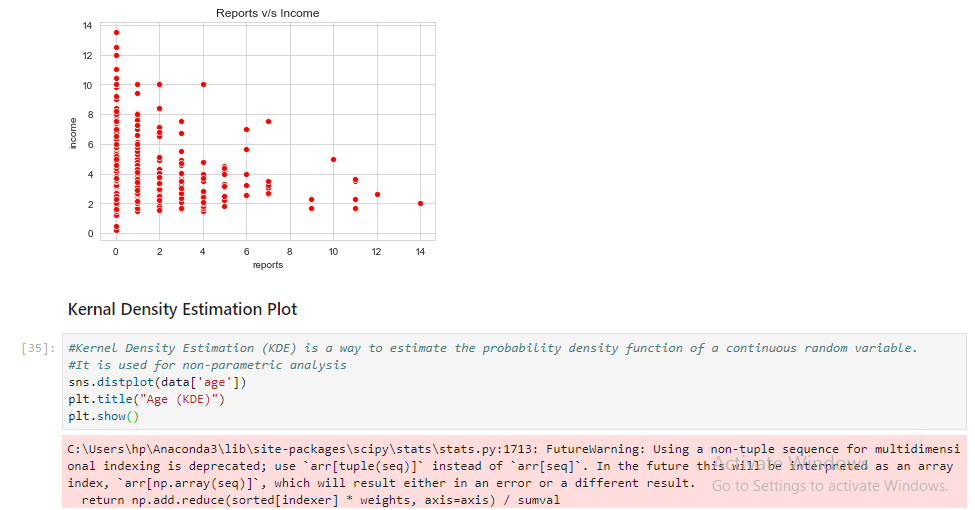
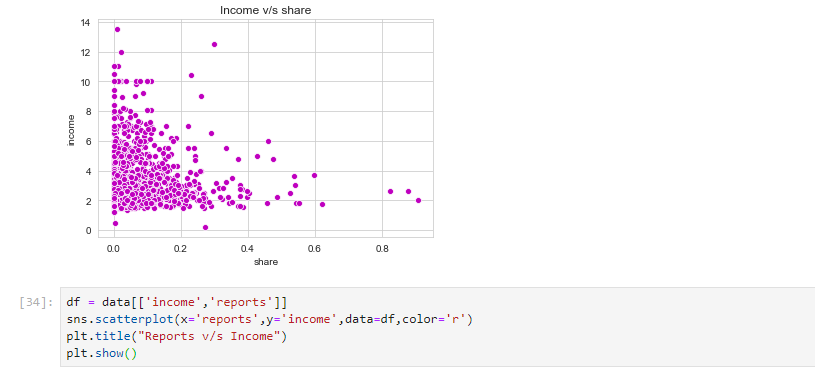
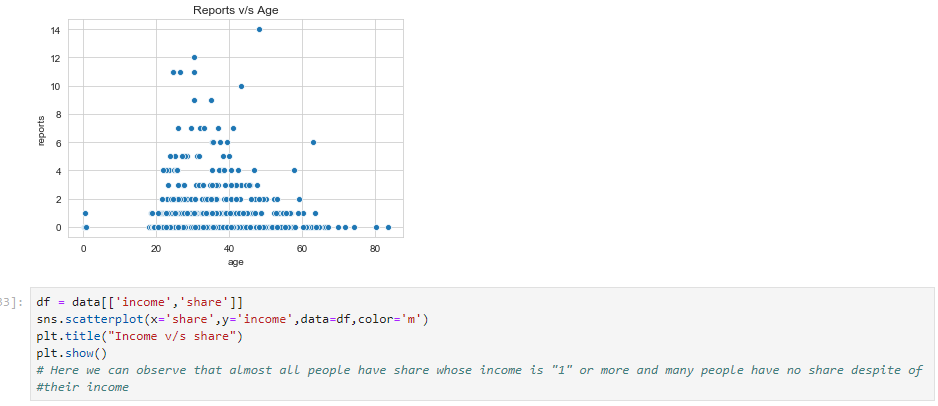
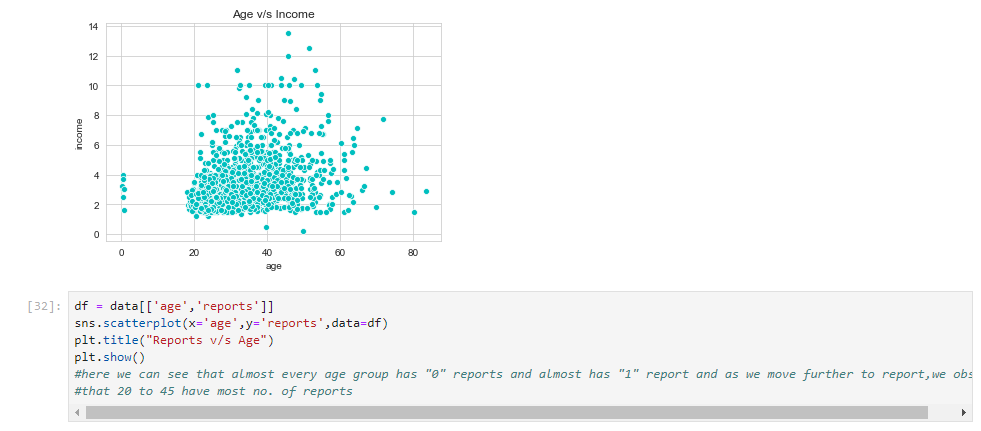
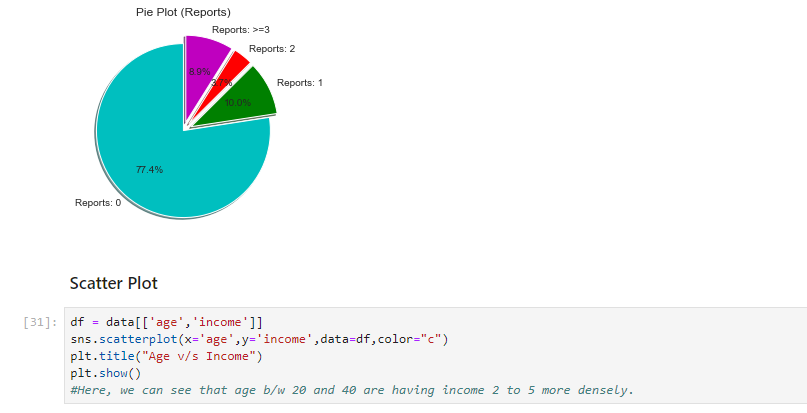
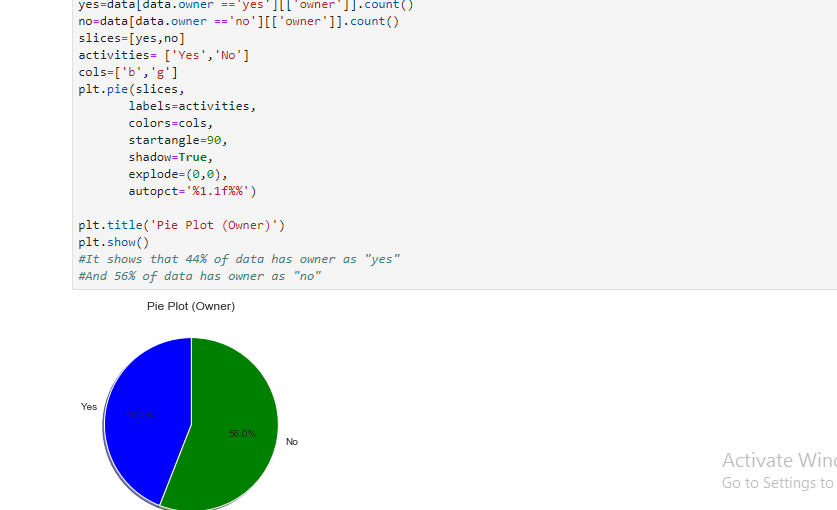
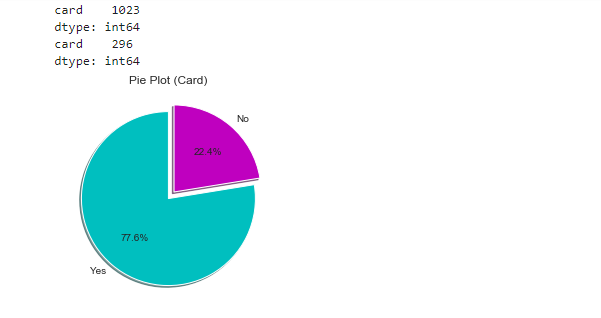
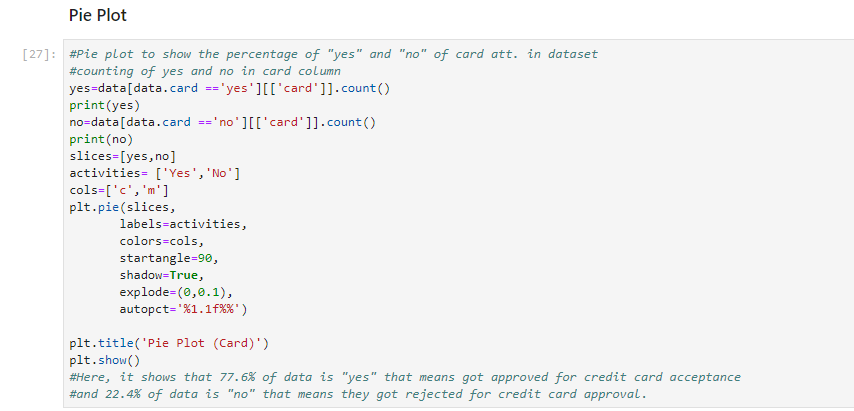
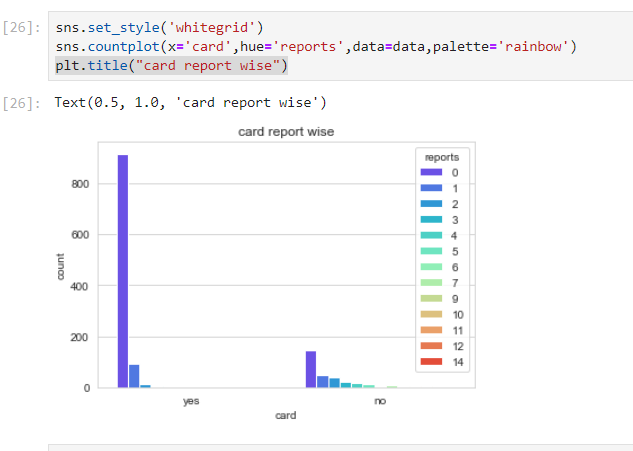
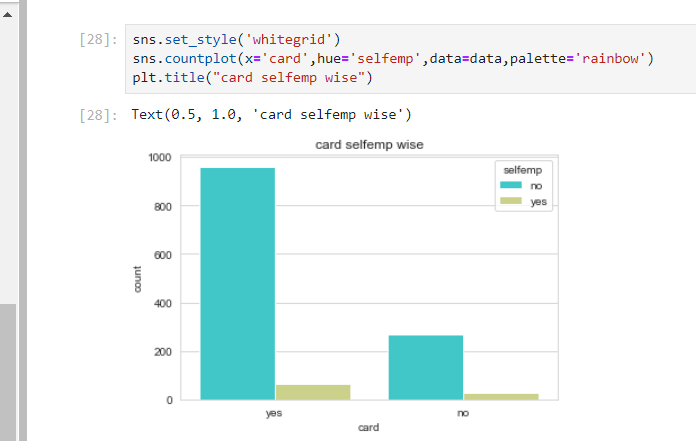
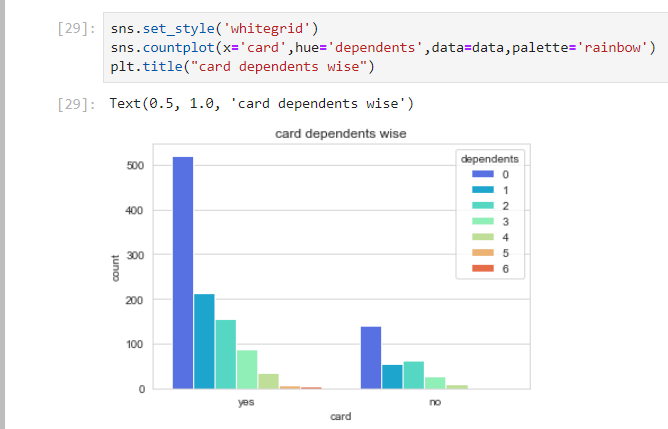
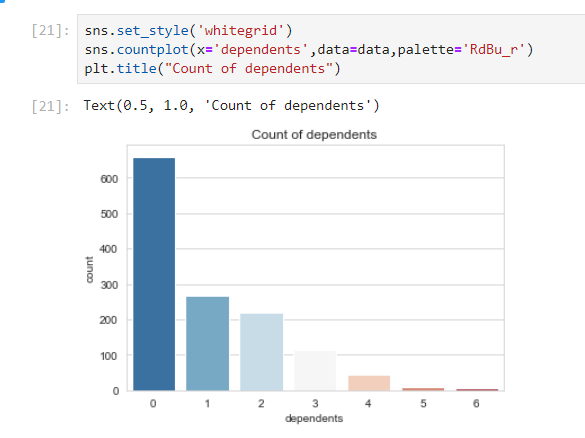
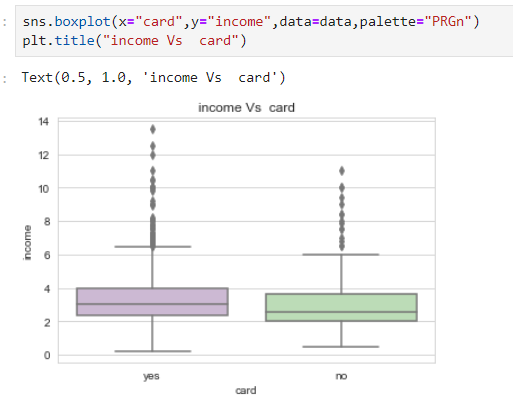
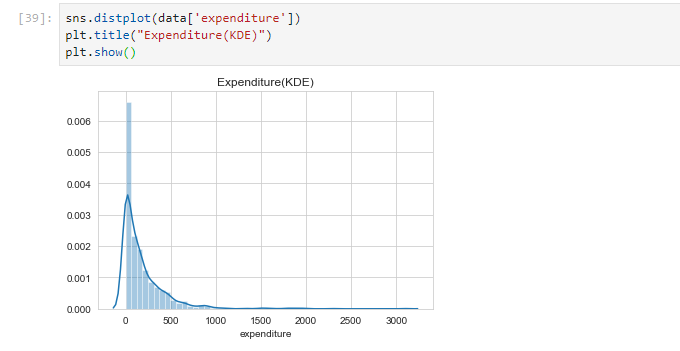
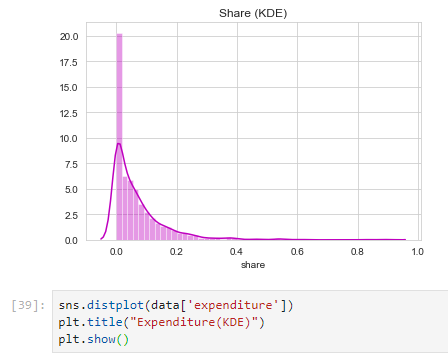
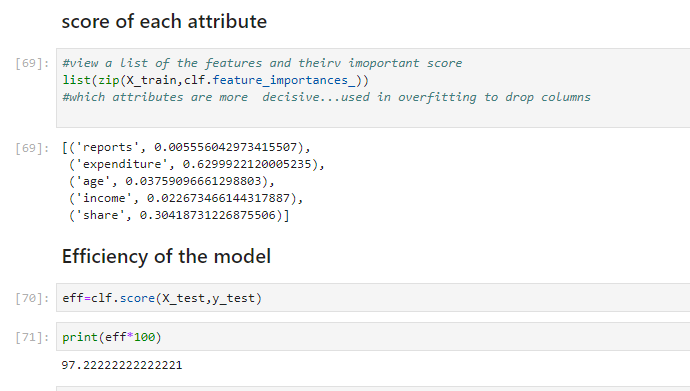
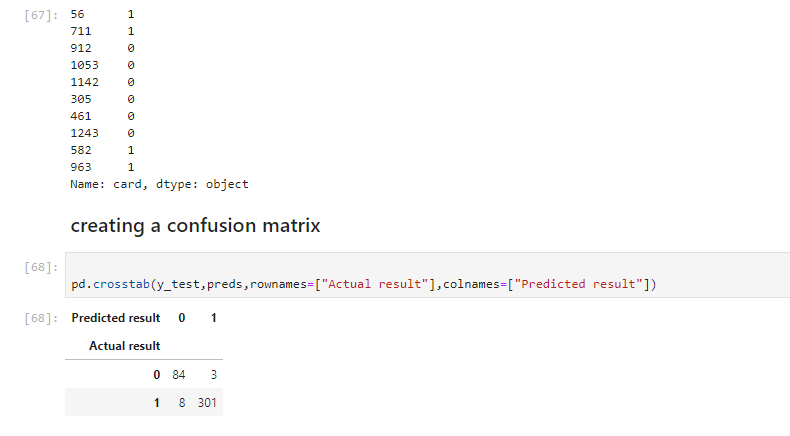
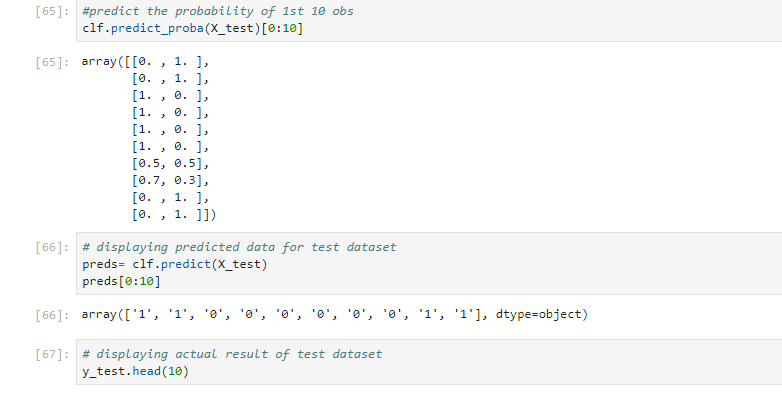
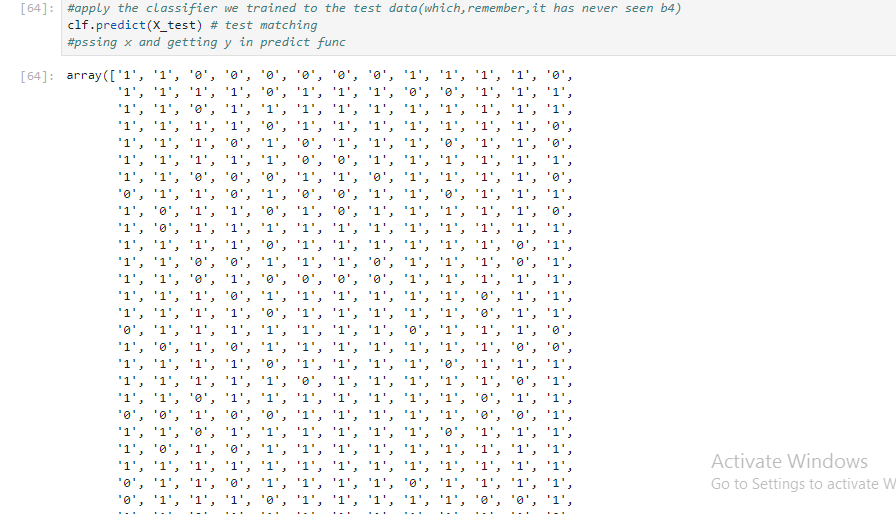
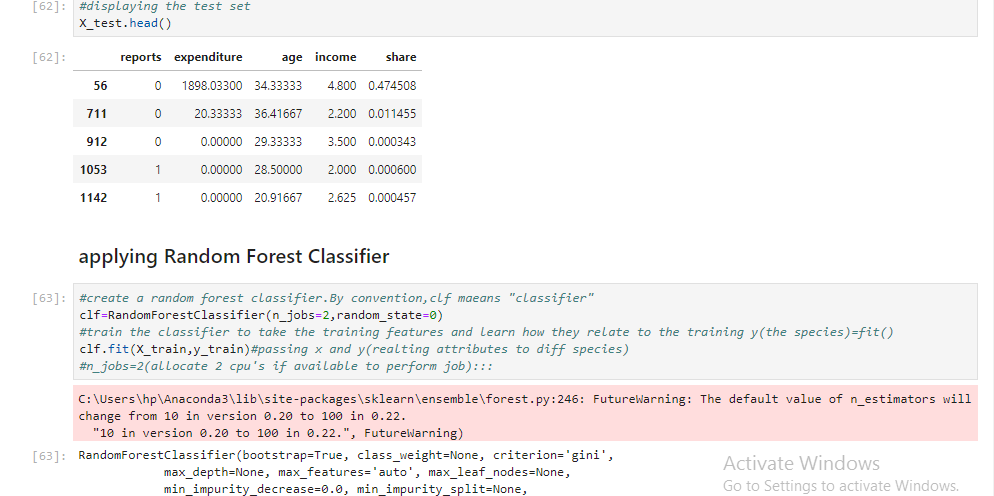
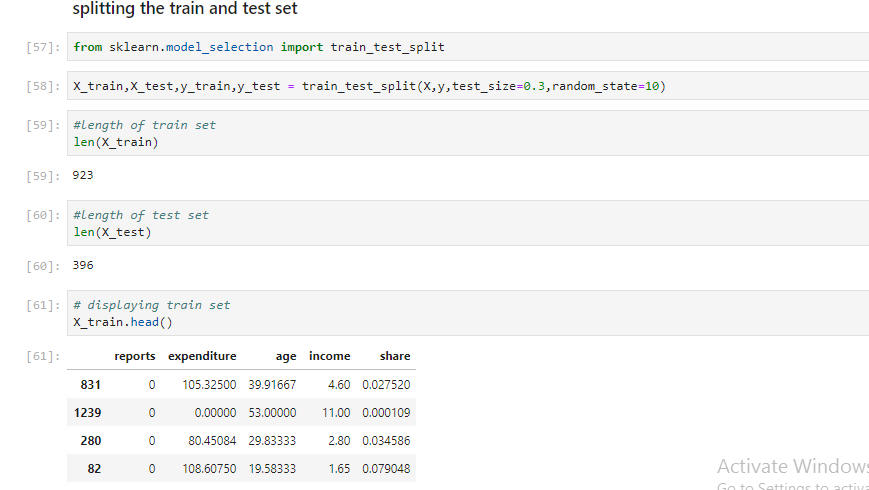
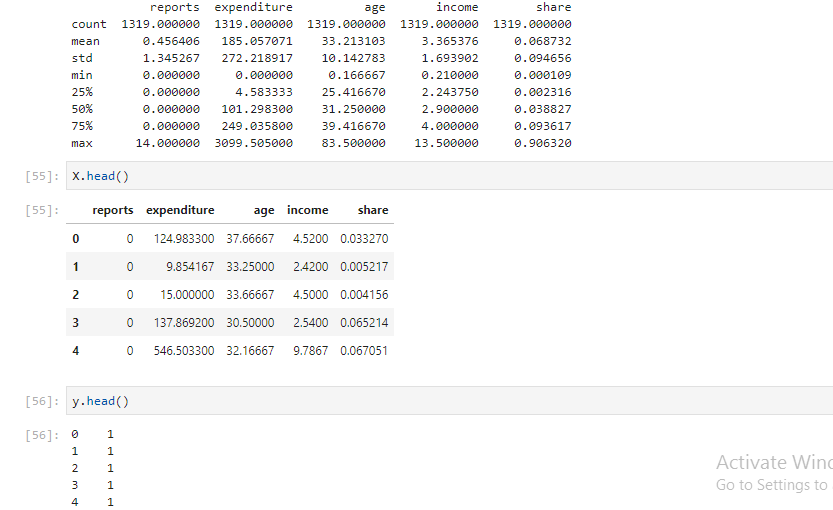
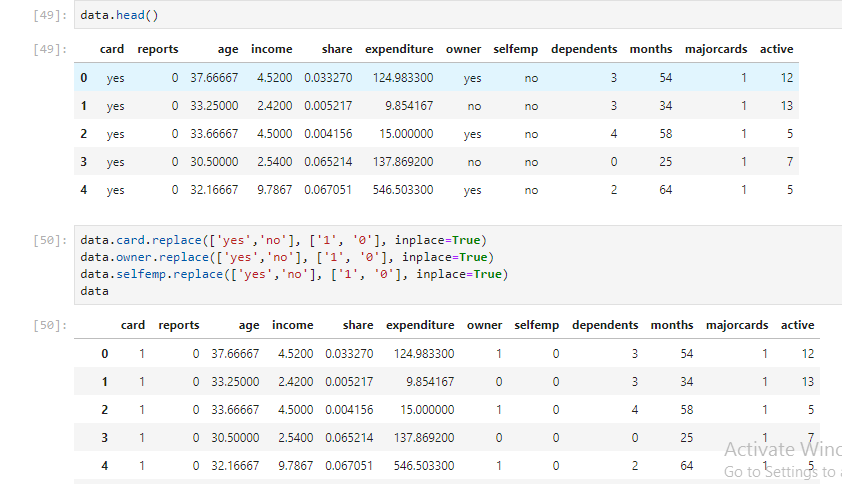
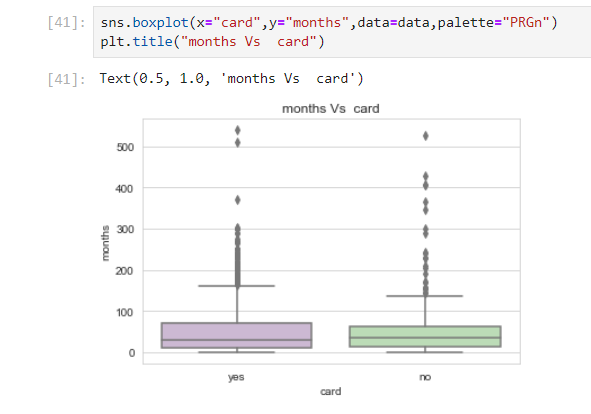
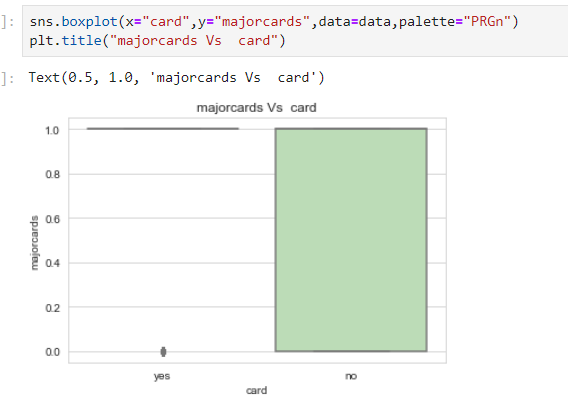
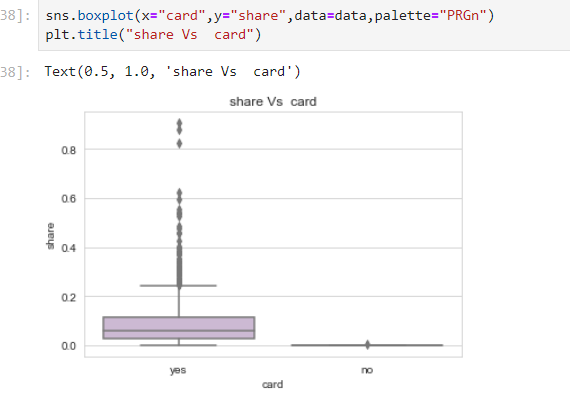
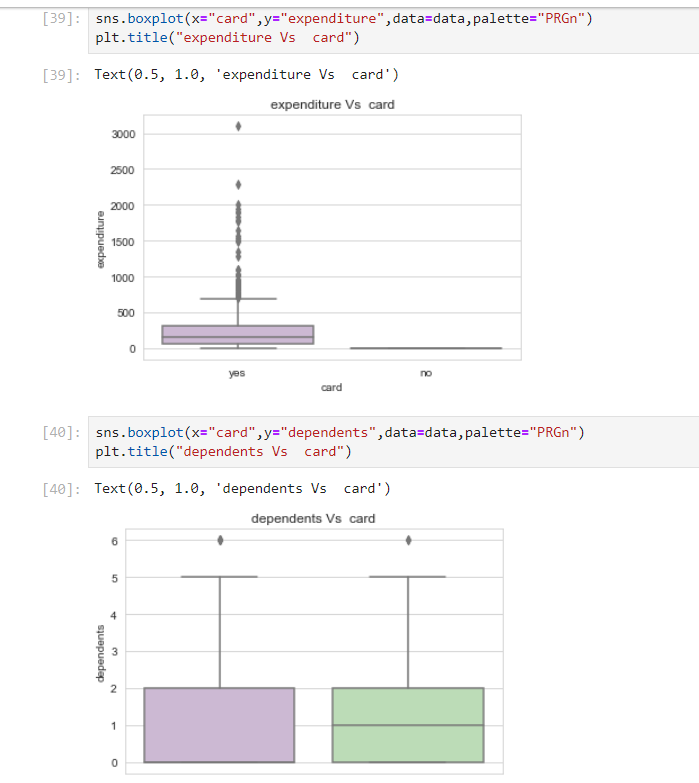






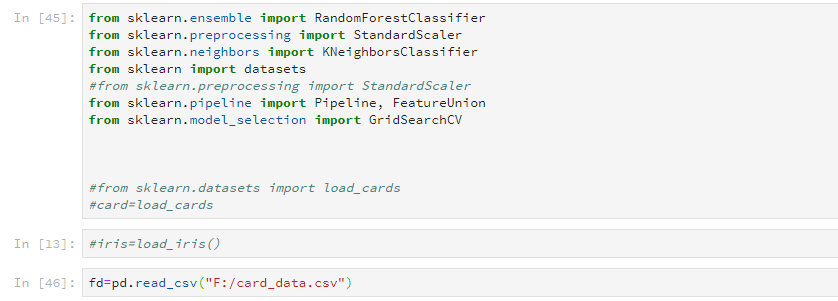


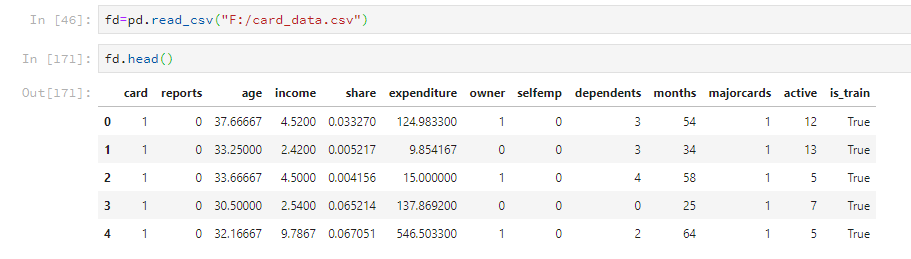


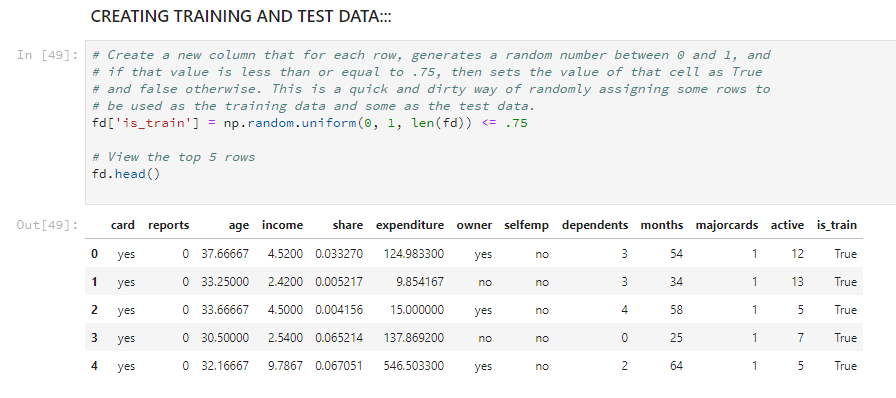
       

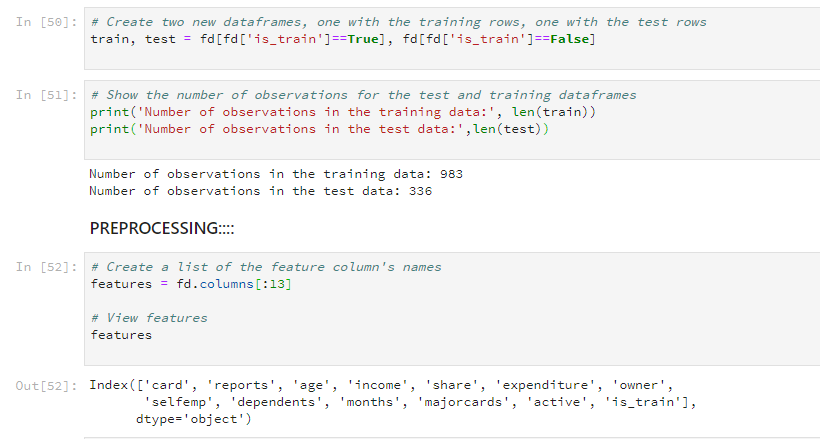
Implementation via KNN:

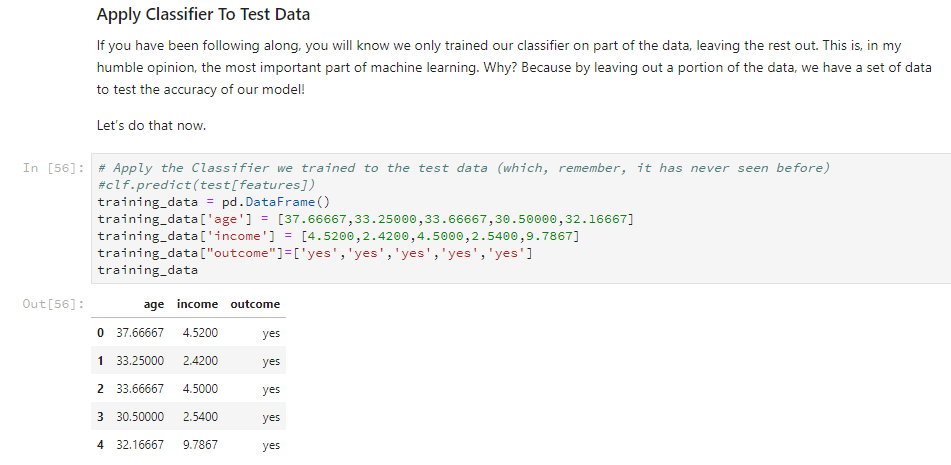


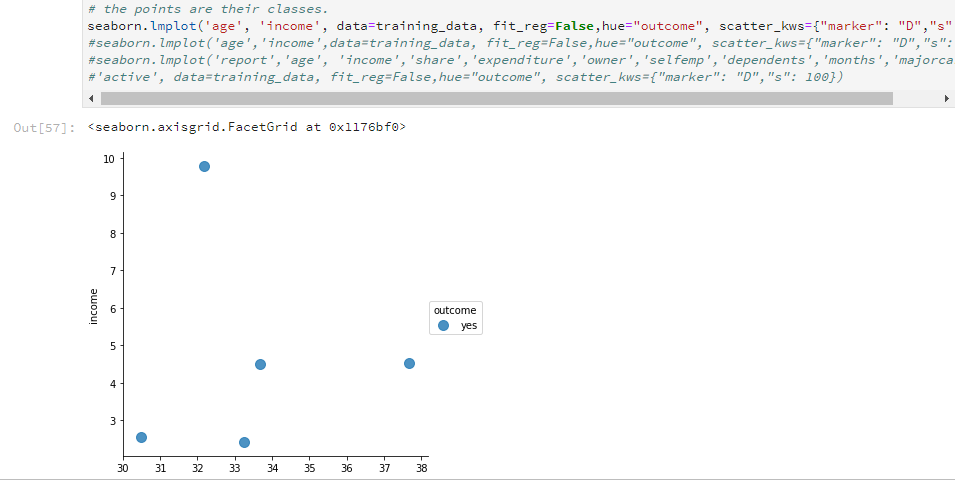


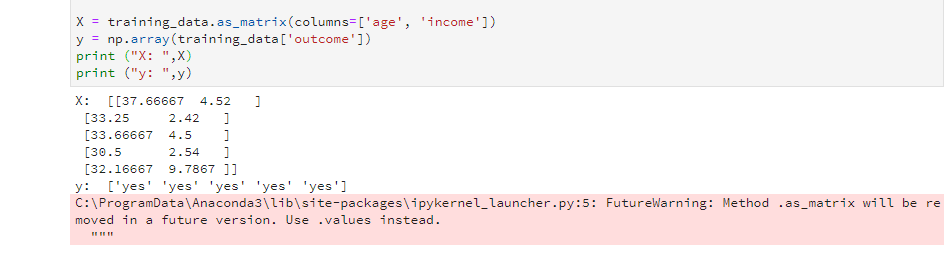


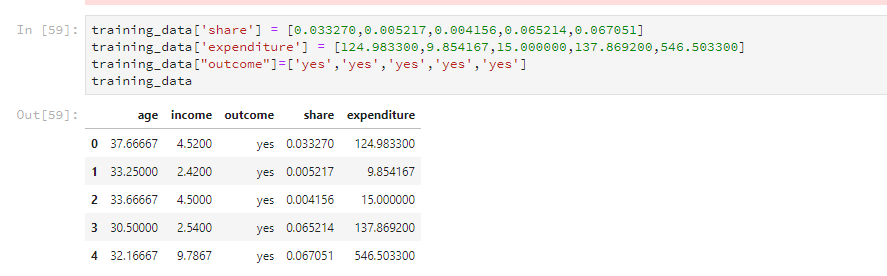




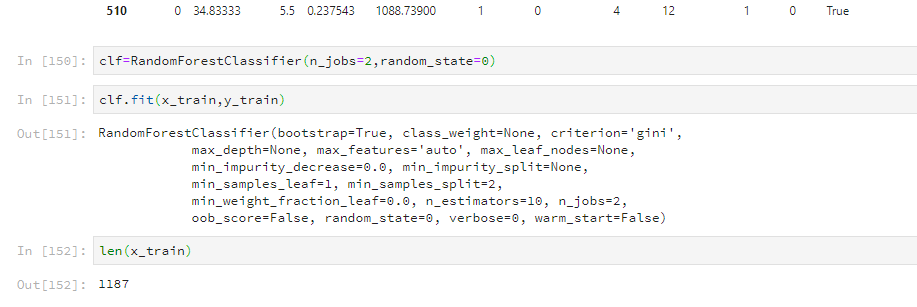


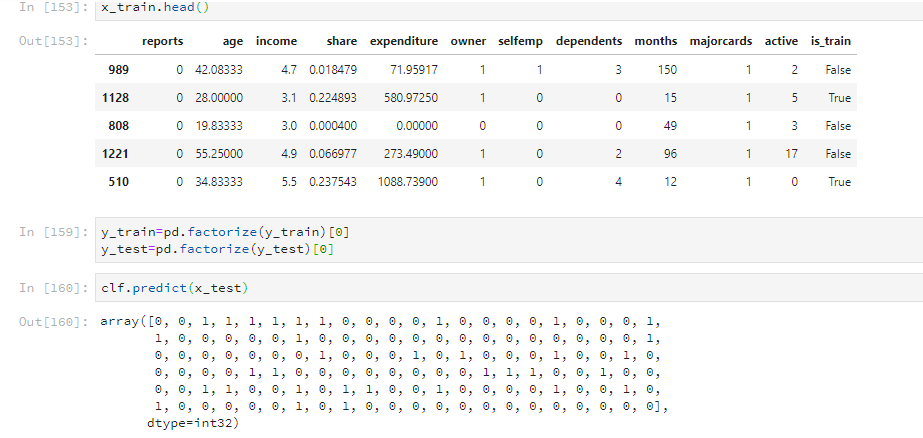


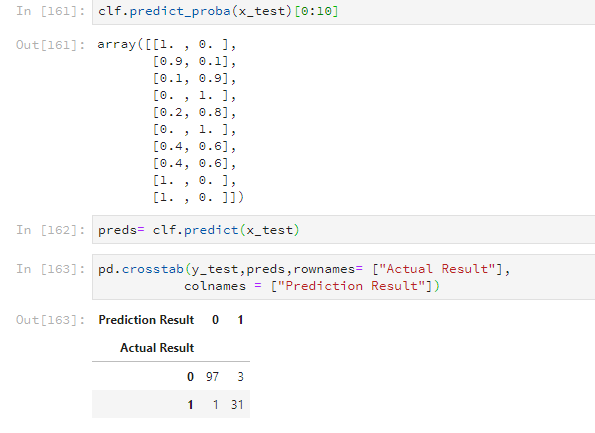


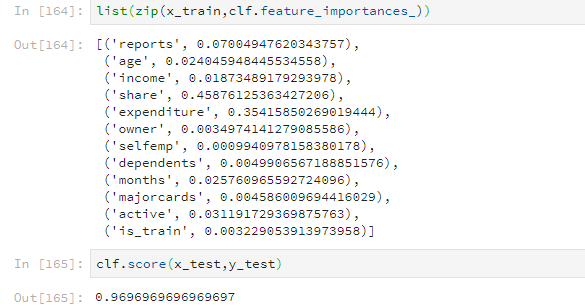




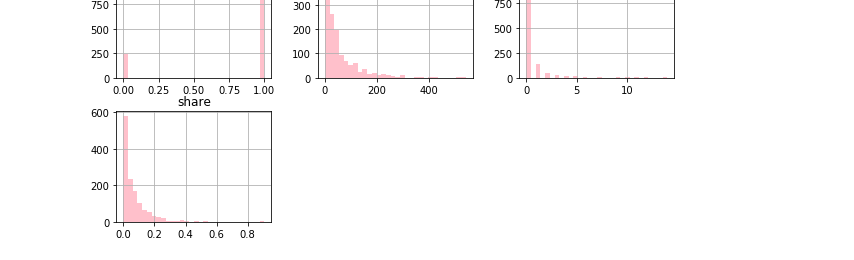


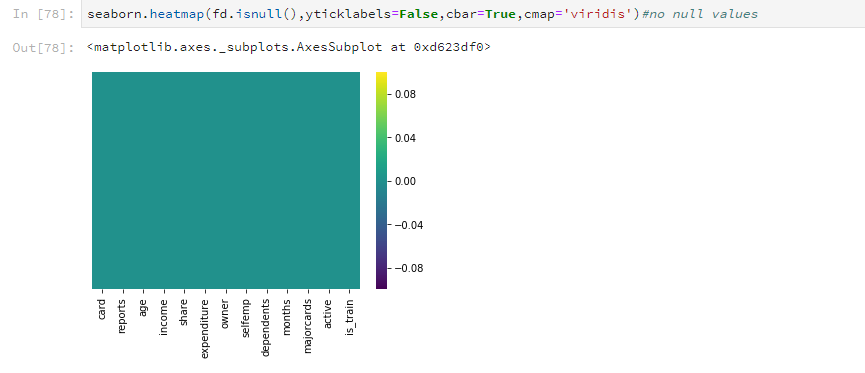


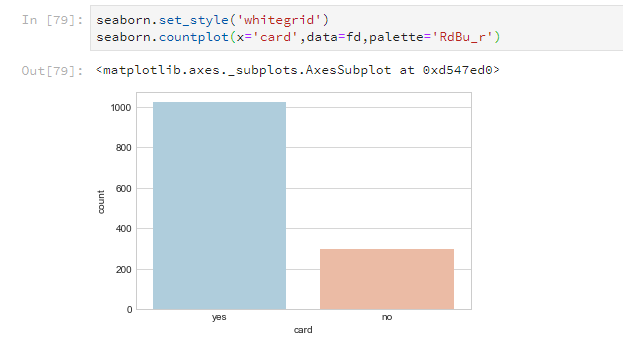


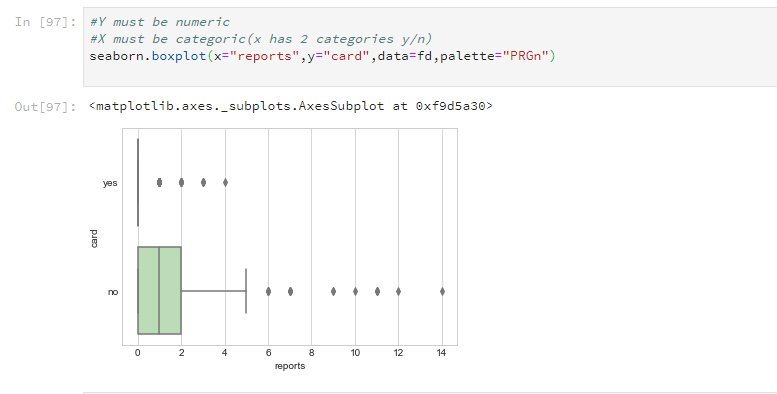


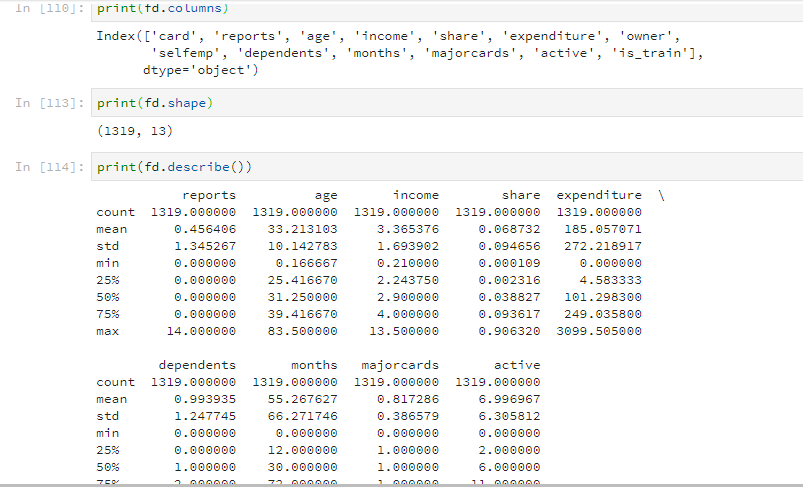


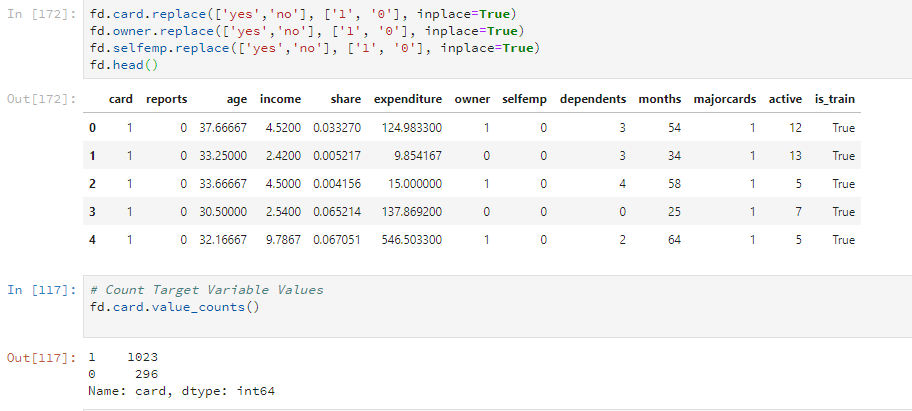






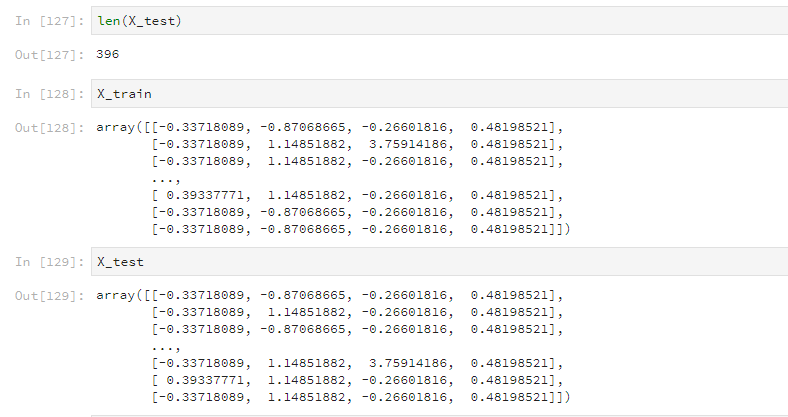




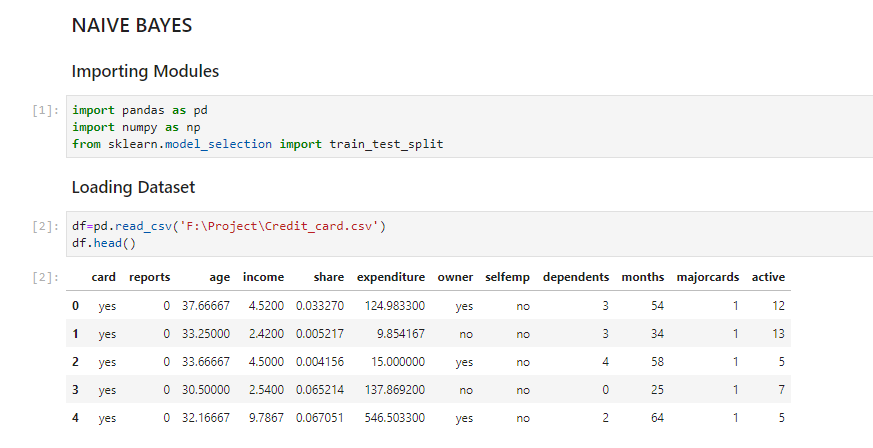


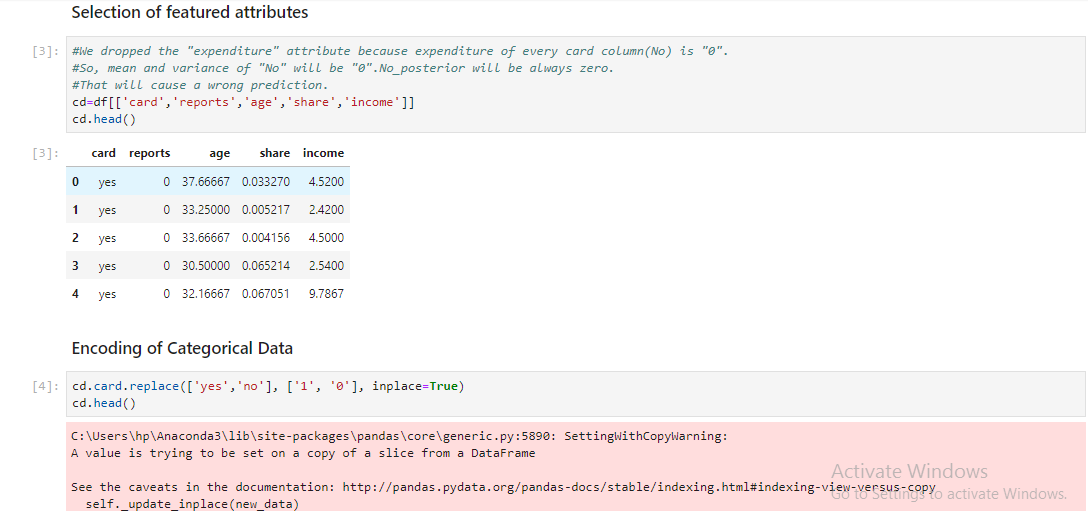


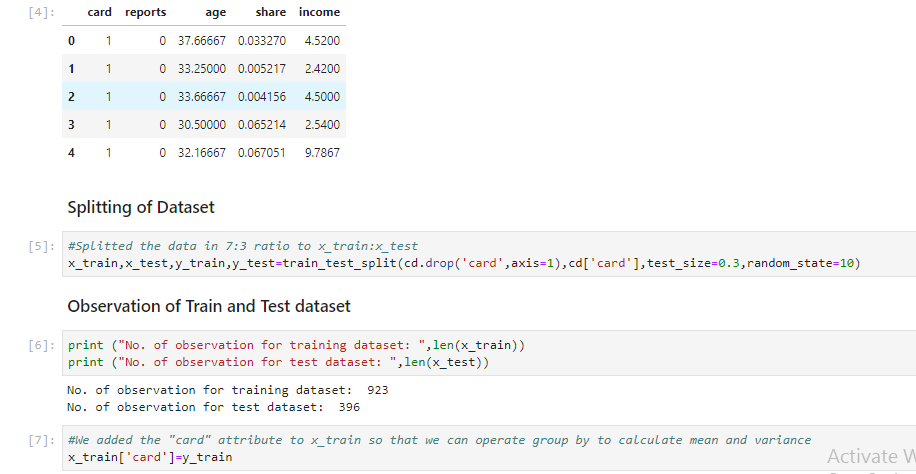




Implementation via Naïve Bayes:

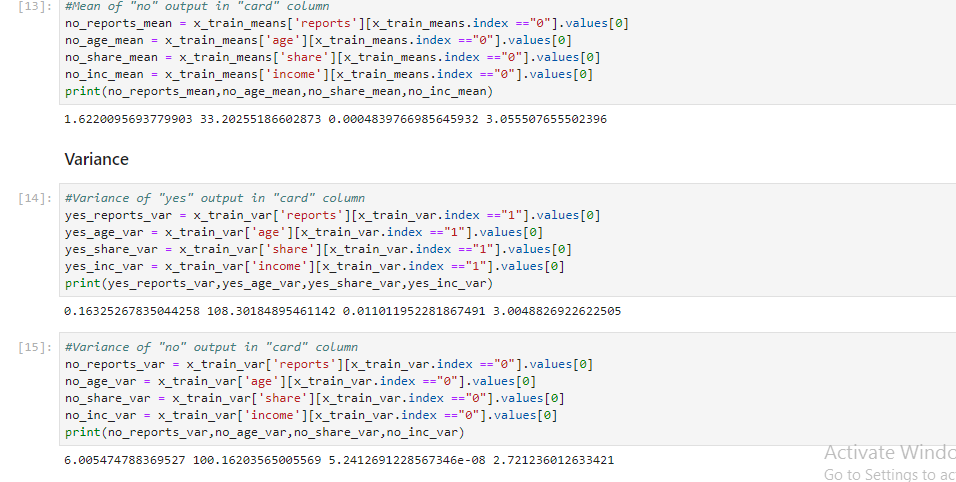




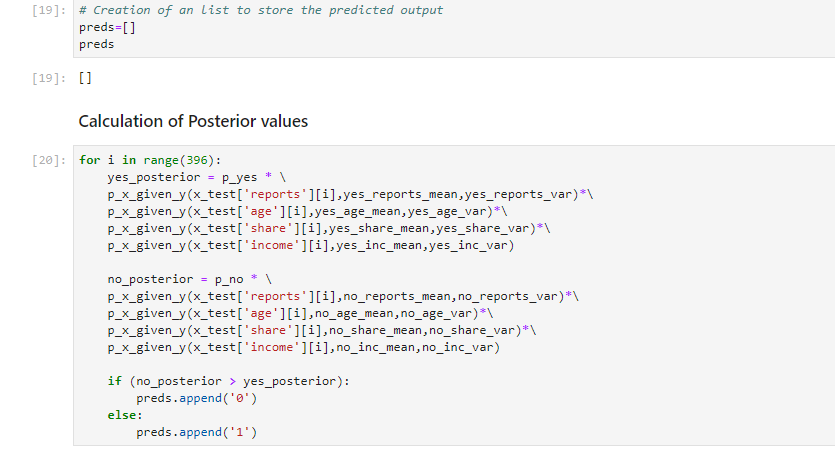


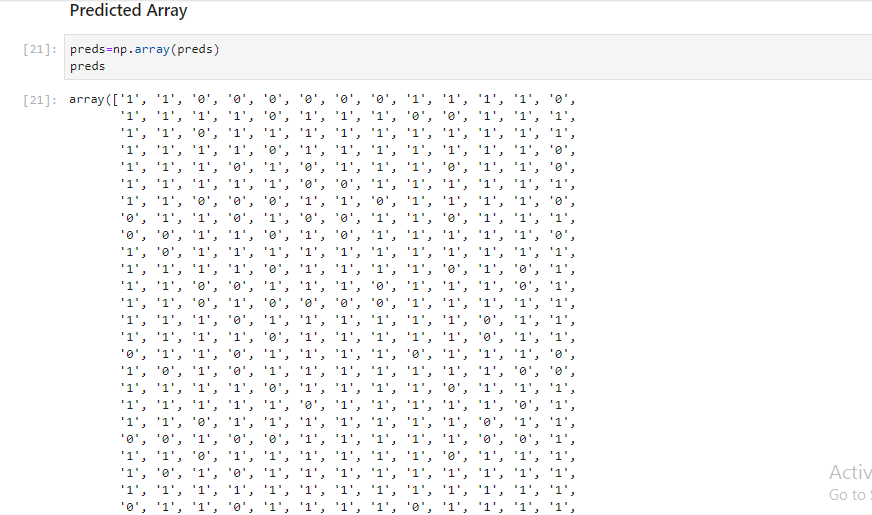


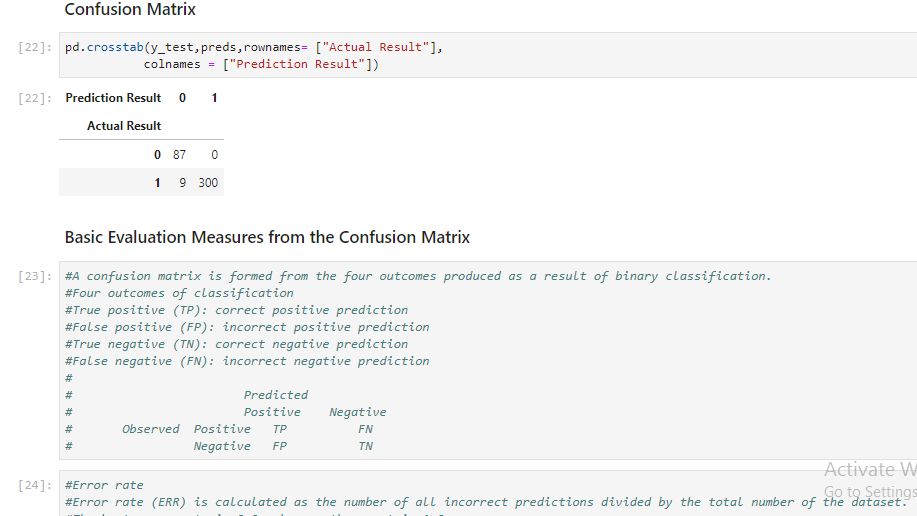


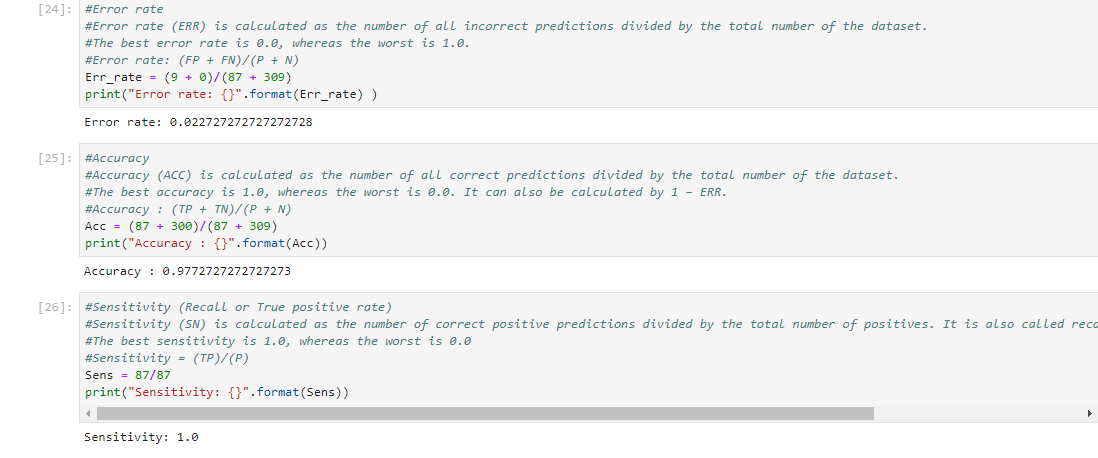


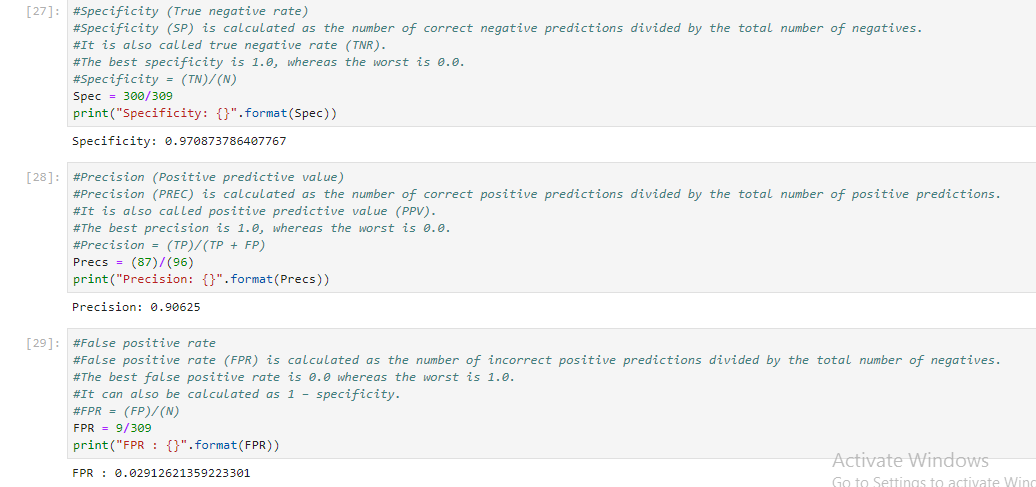




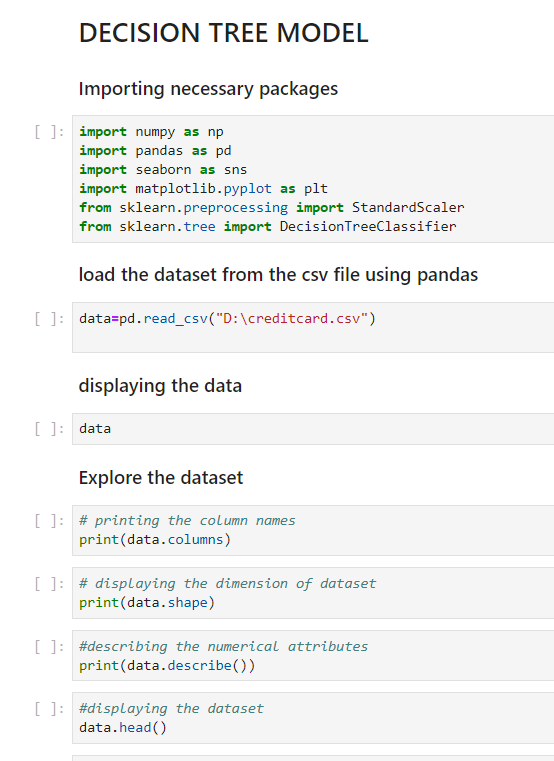


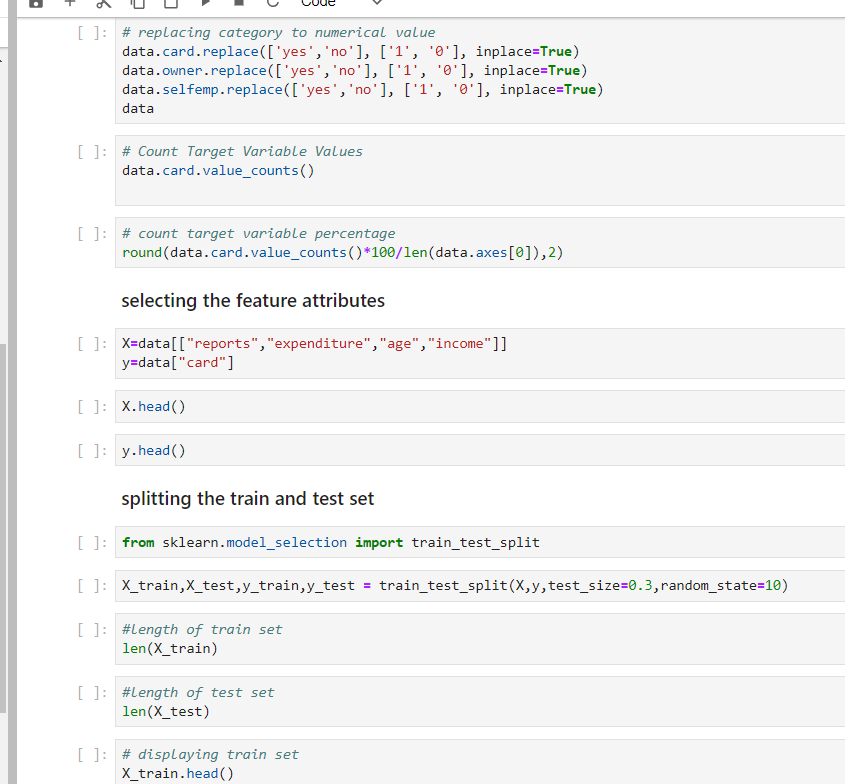


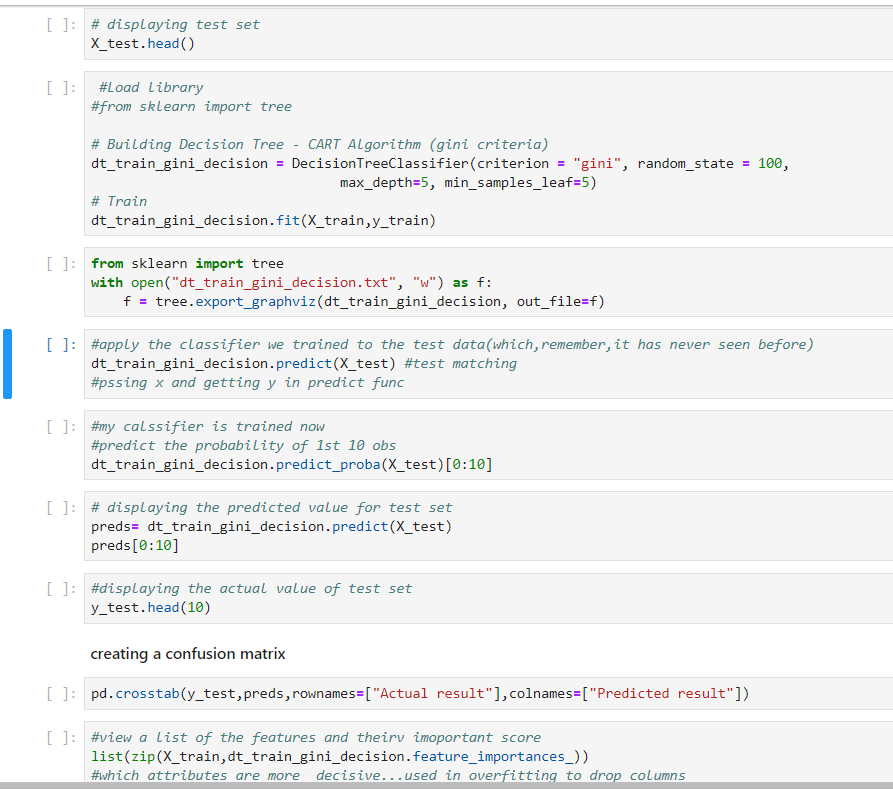




## Implementation via decision tree model:



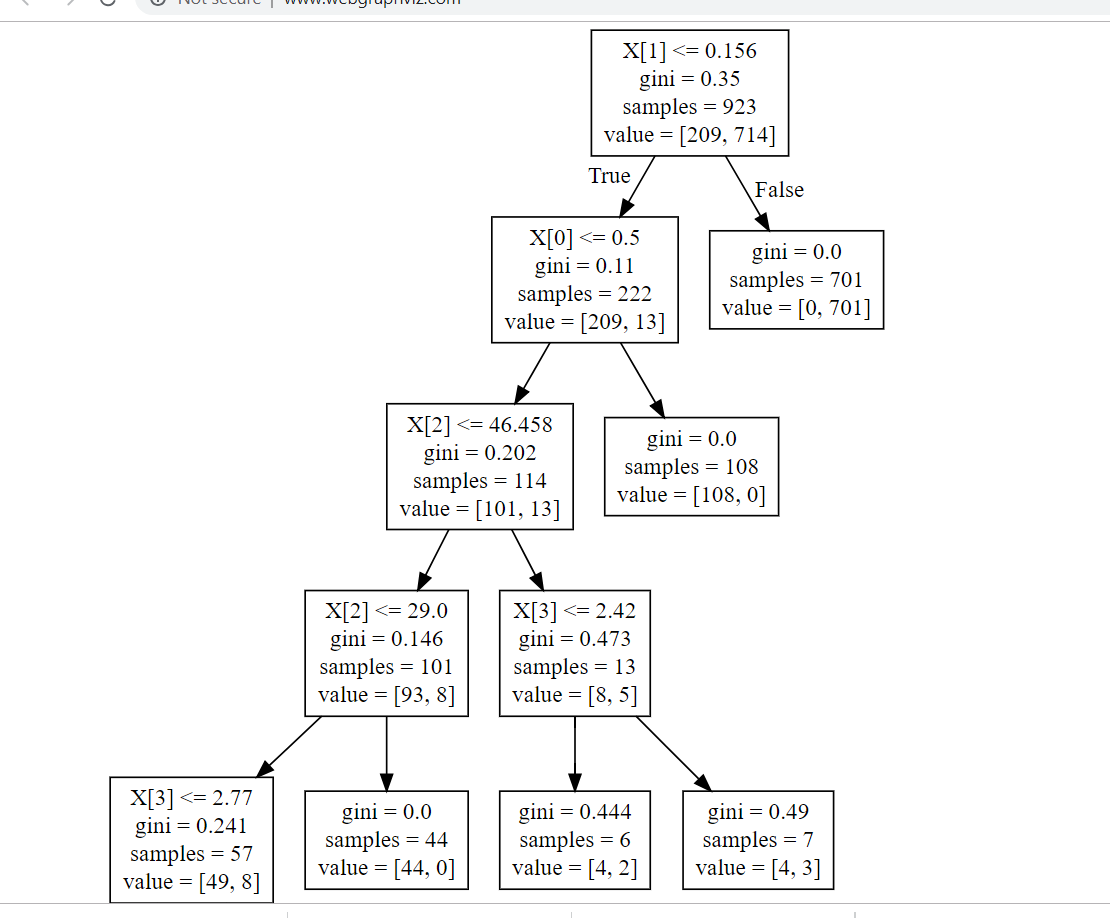




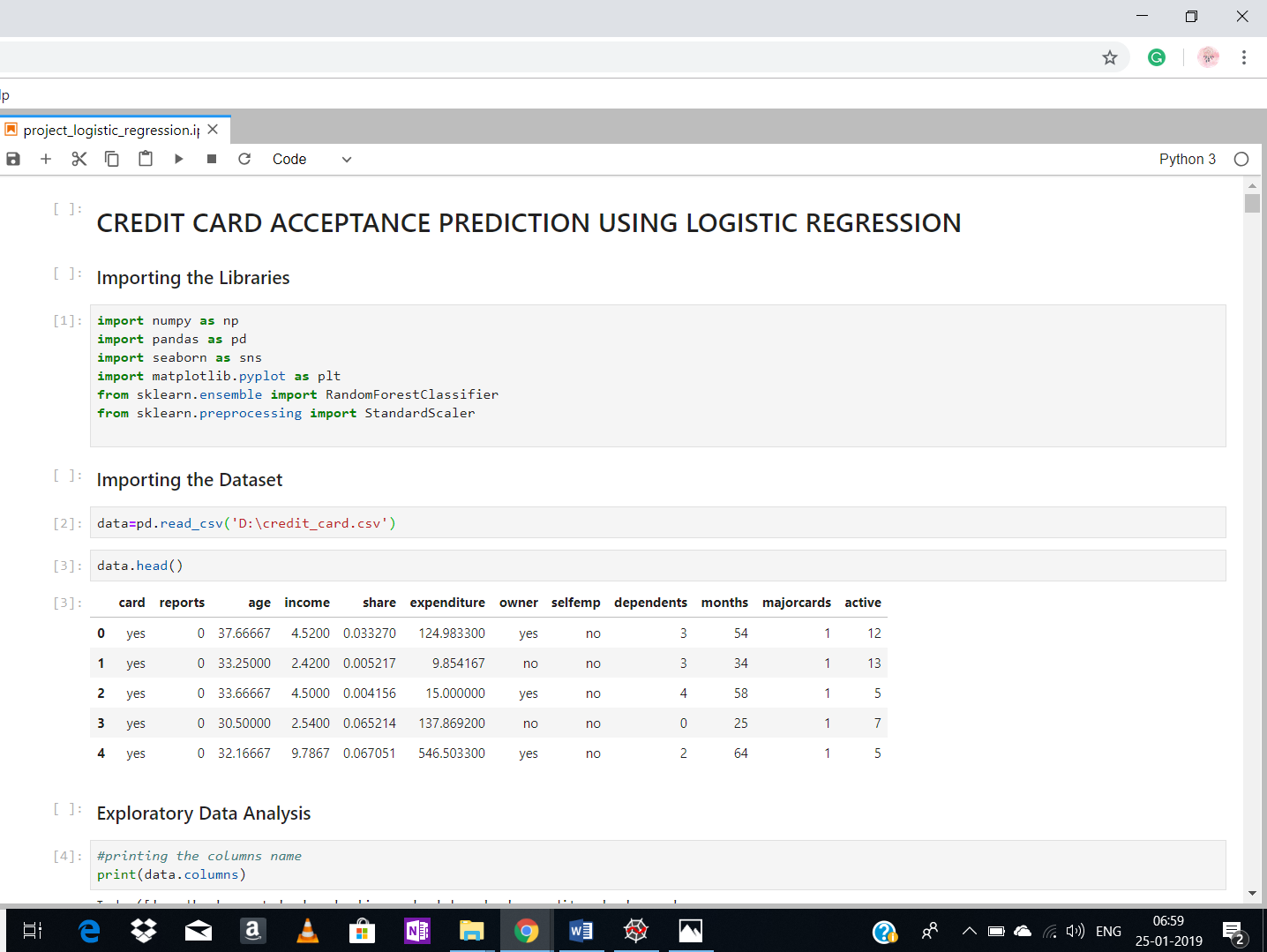


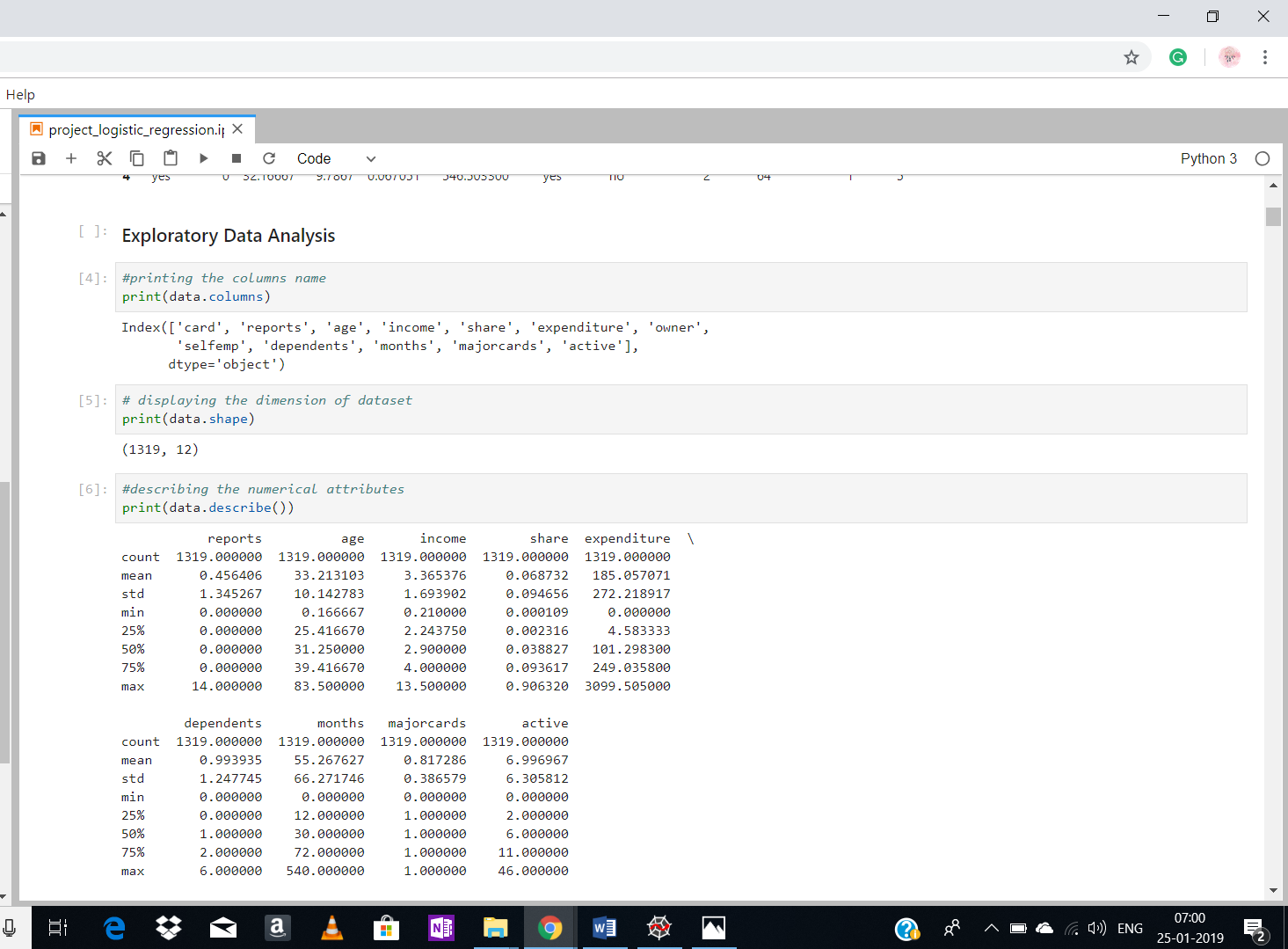


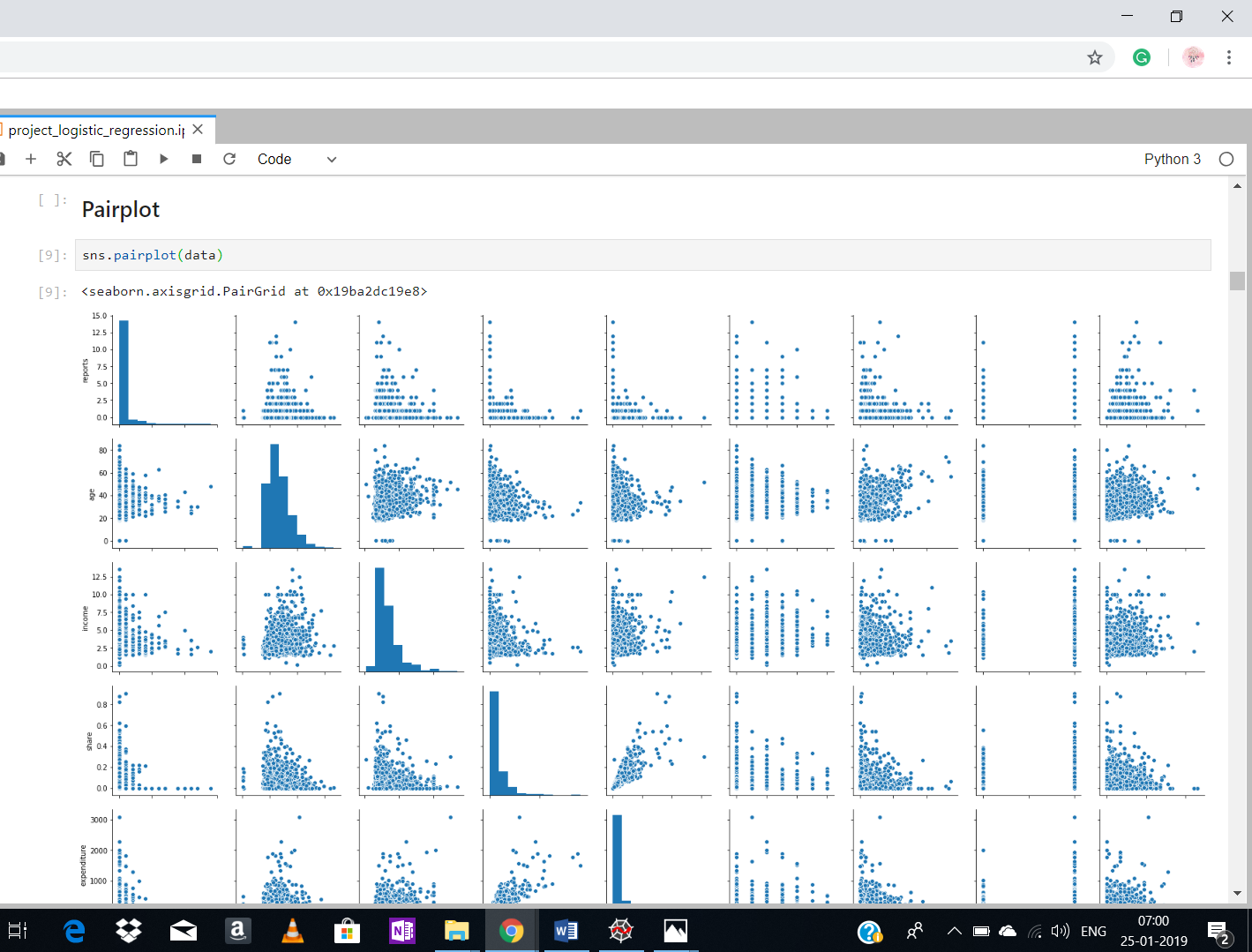
# Decision tree output

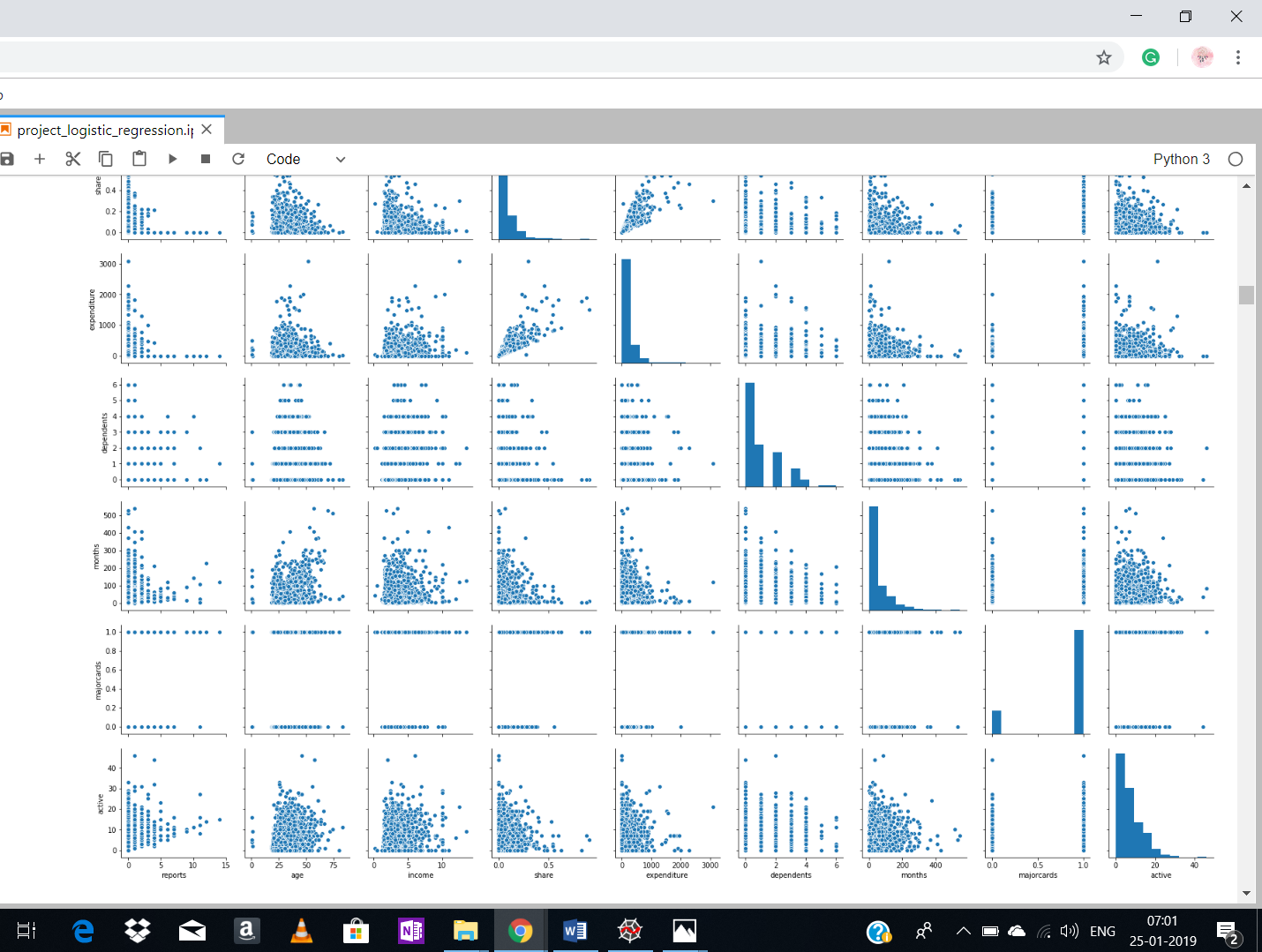


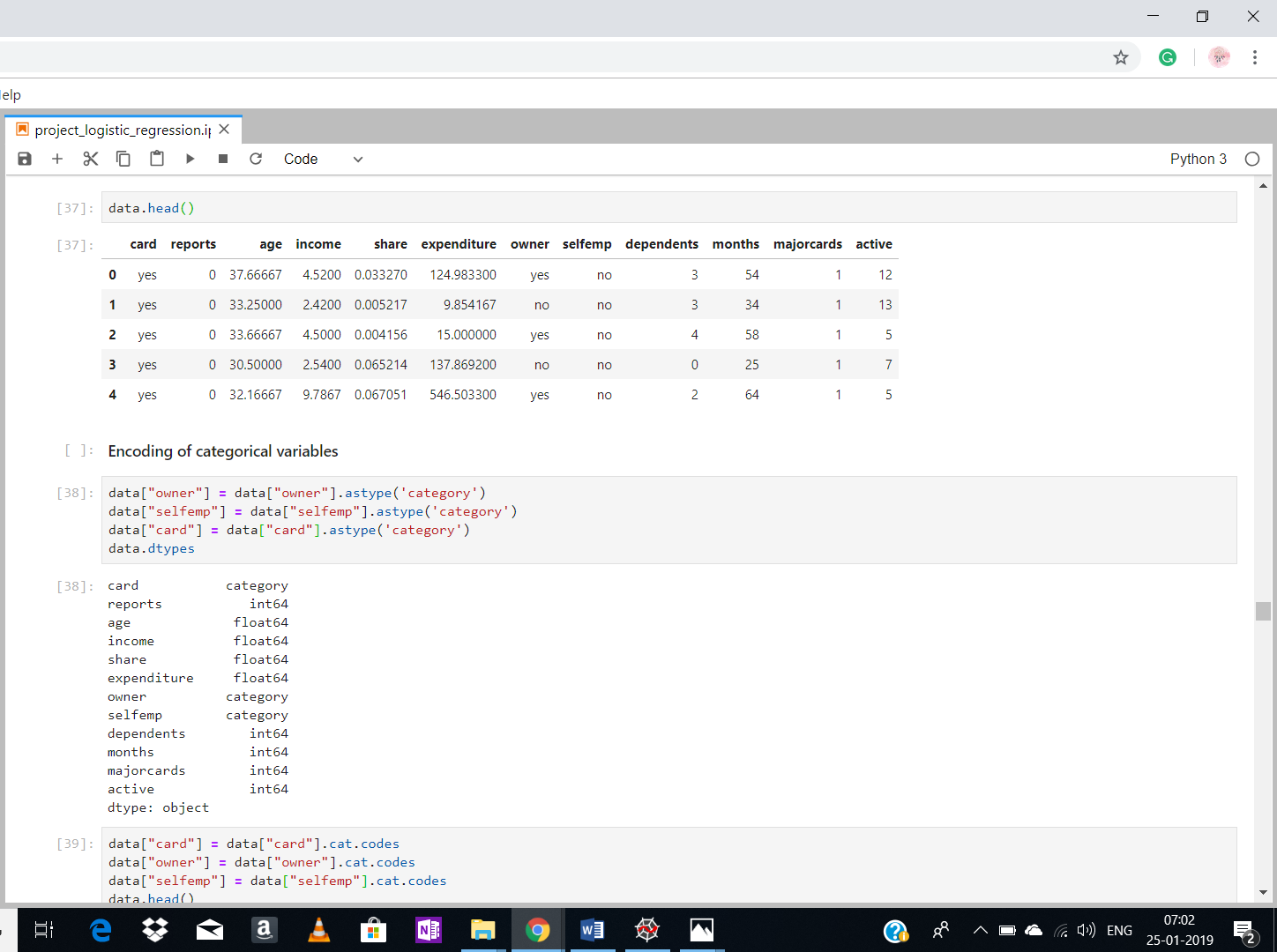
Implementation via Logistic Regression Model:

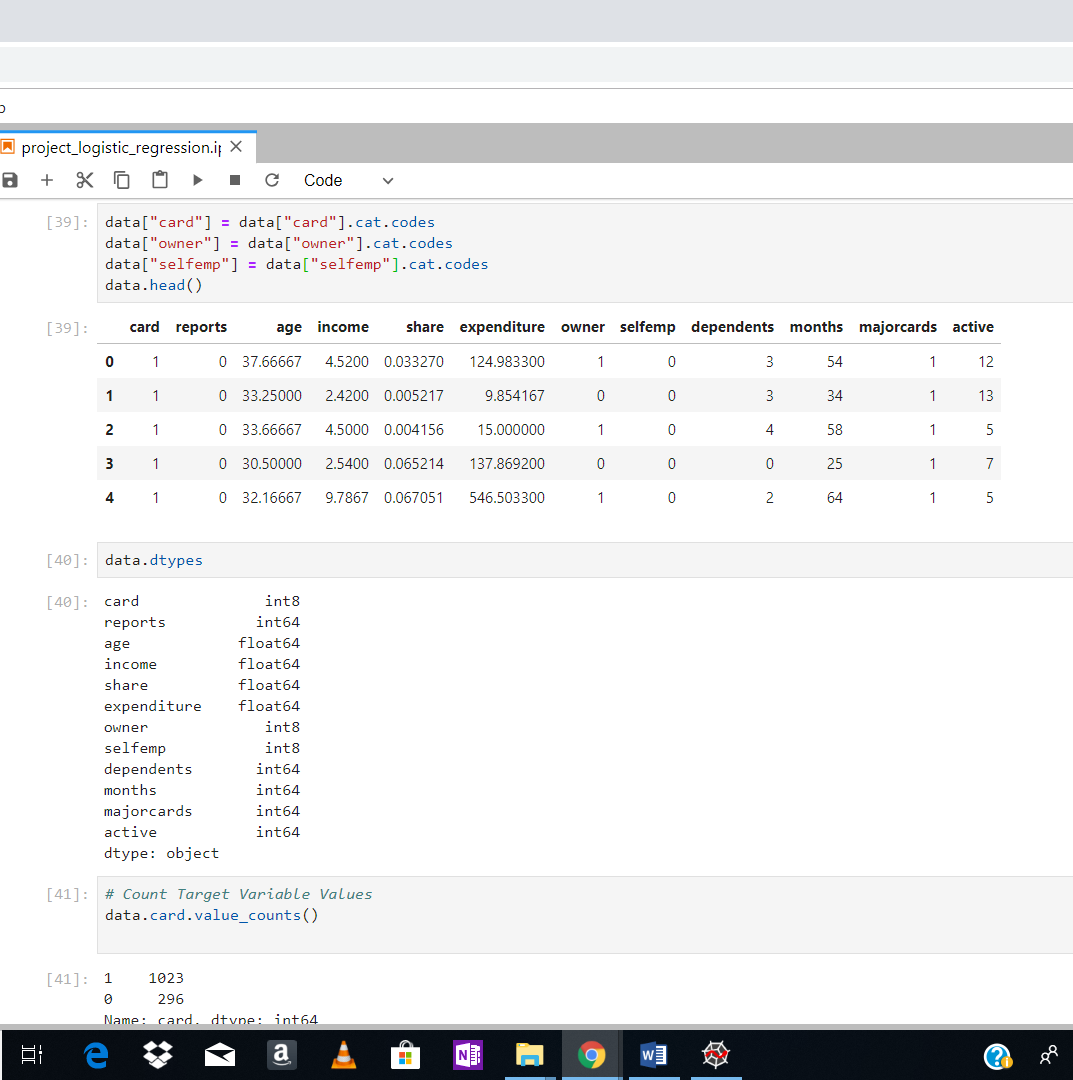
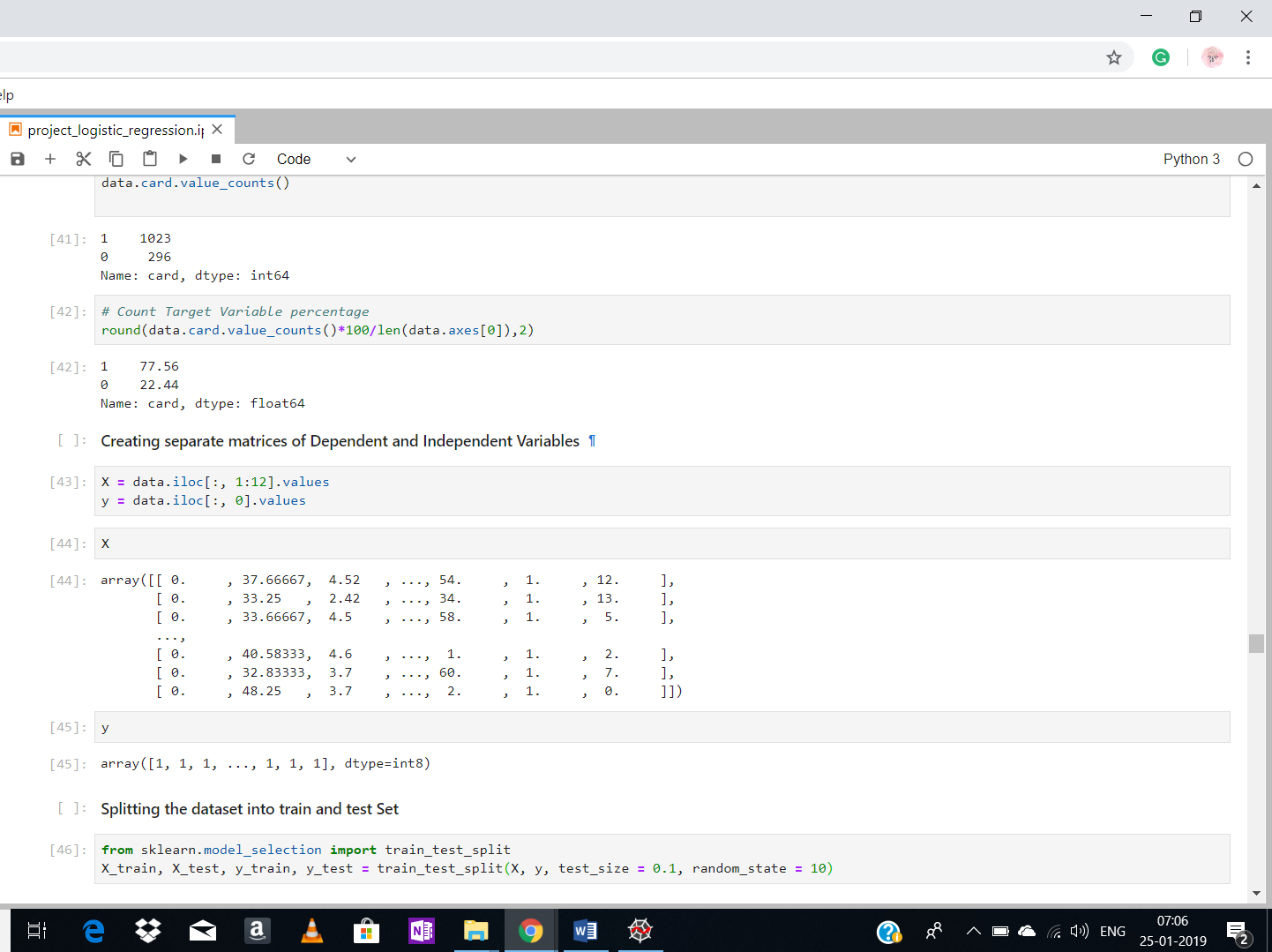


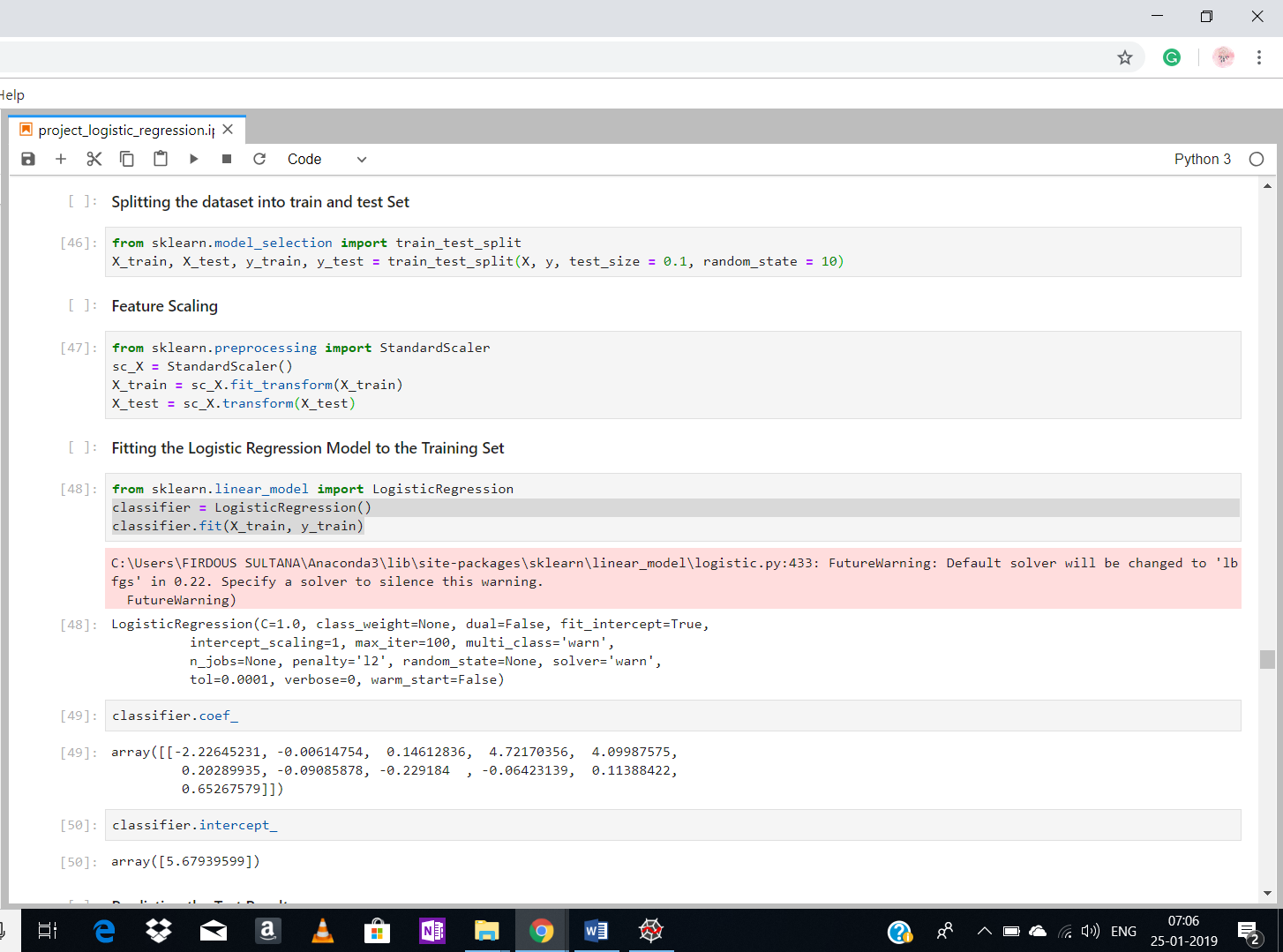


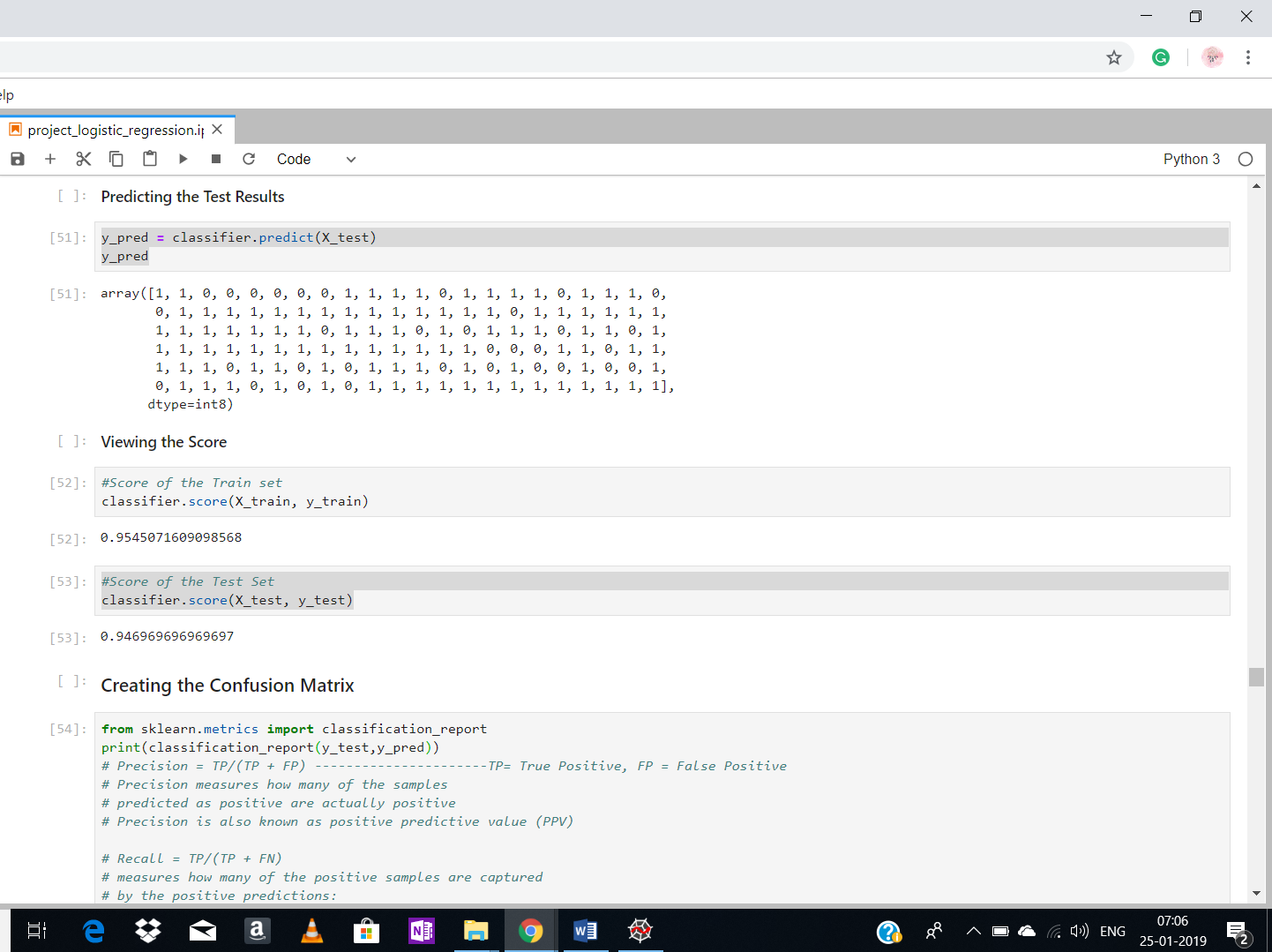




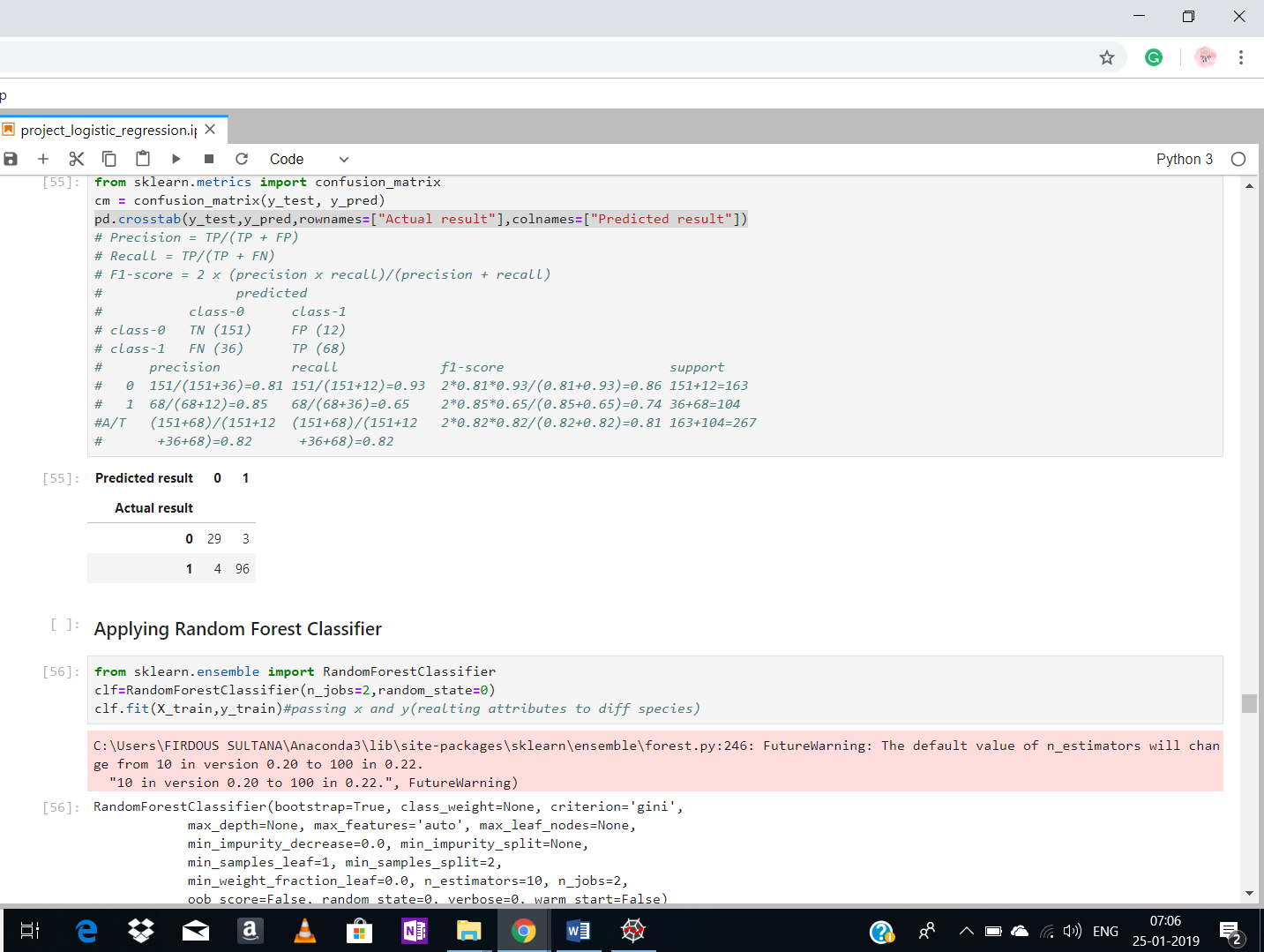


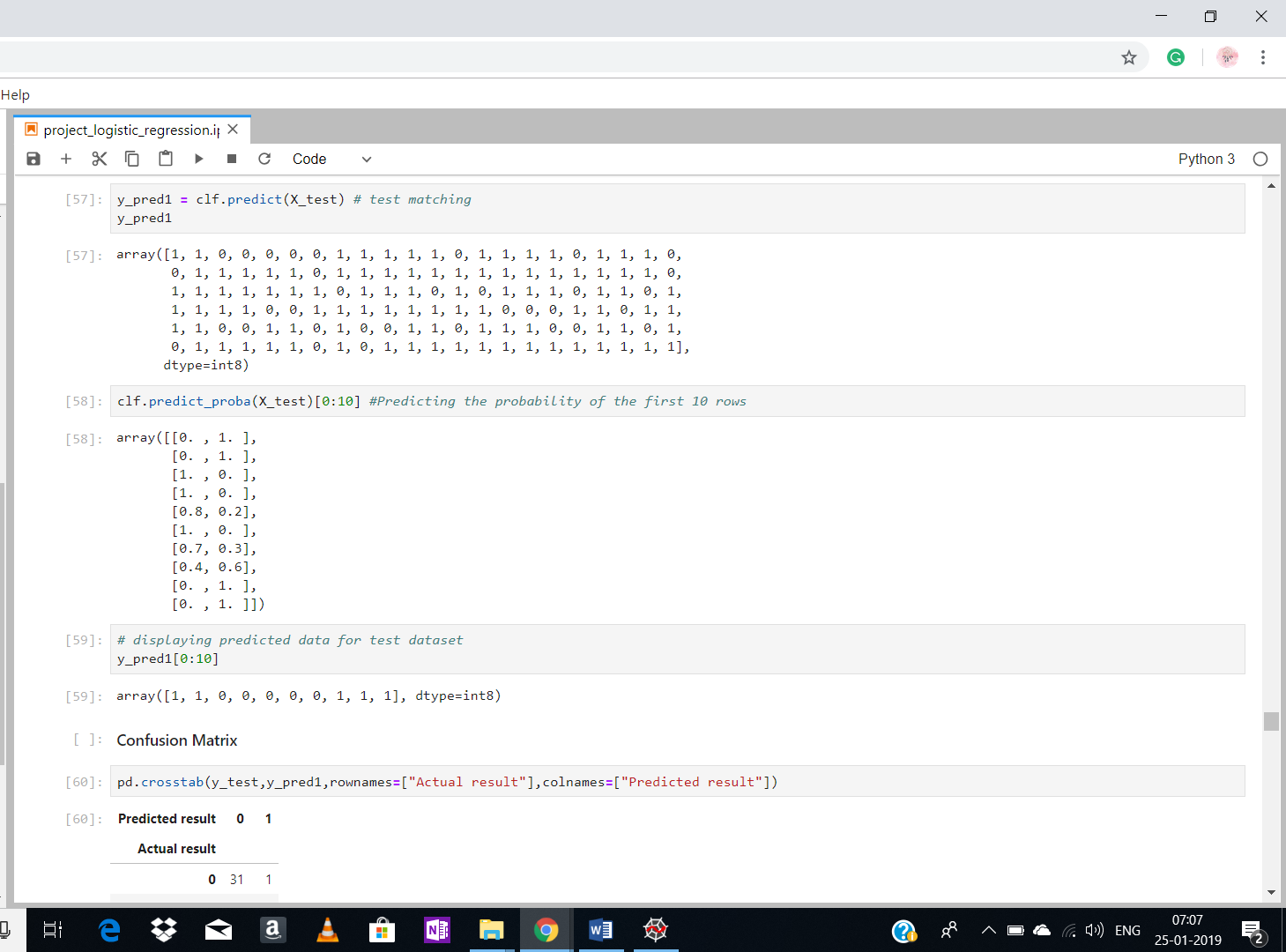


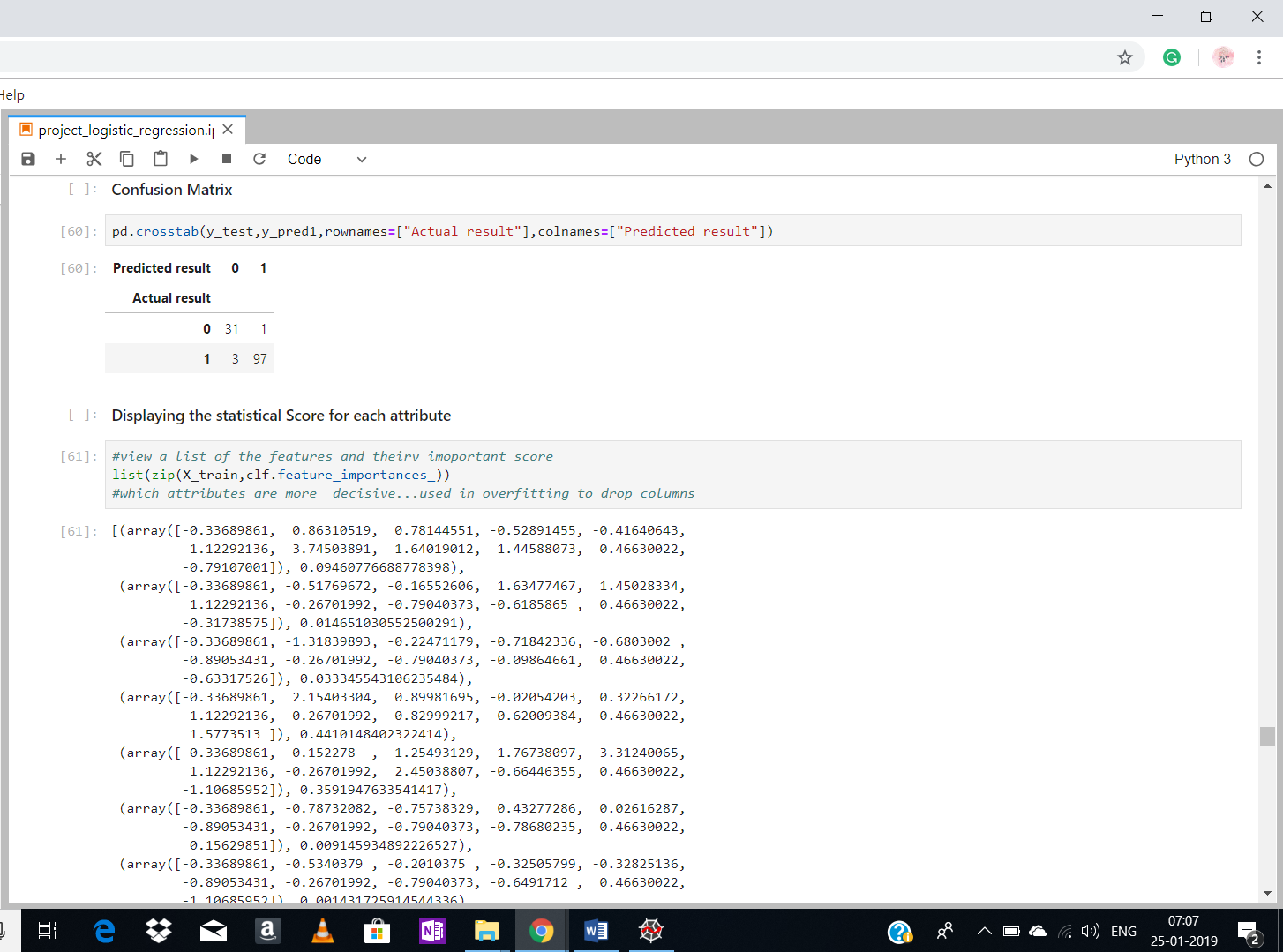


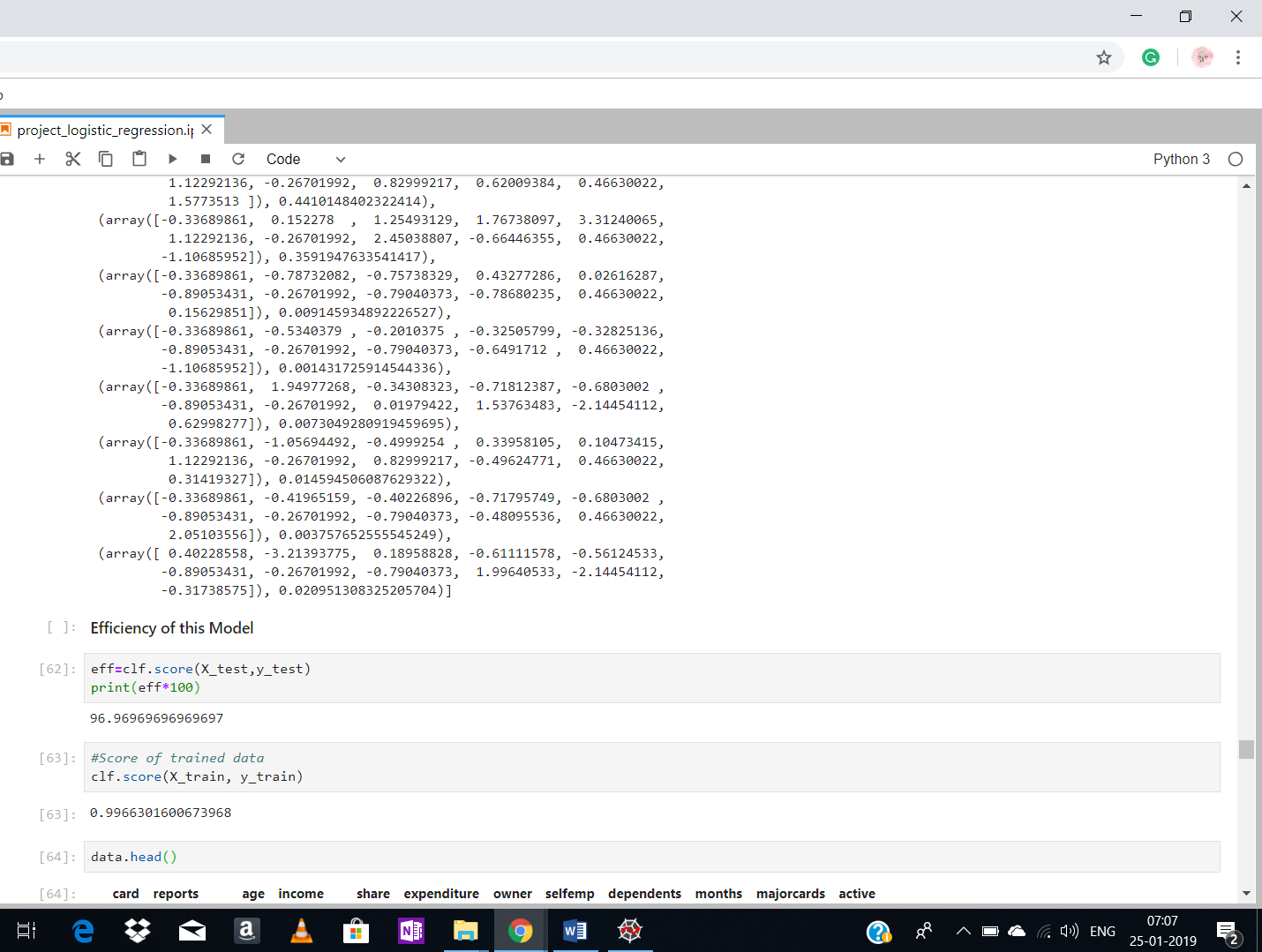


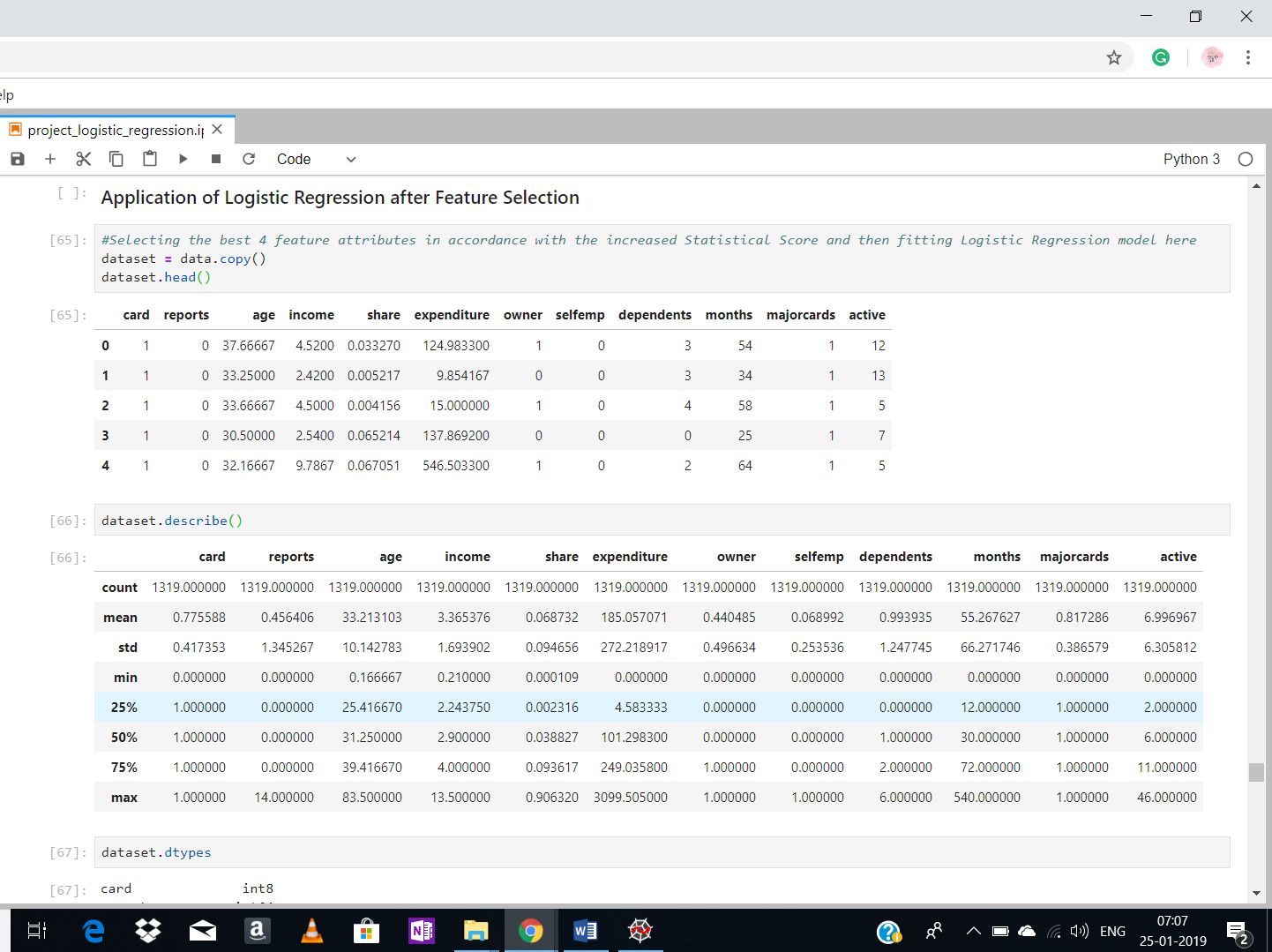


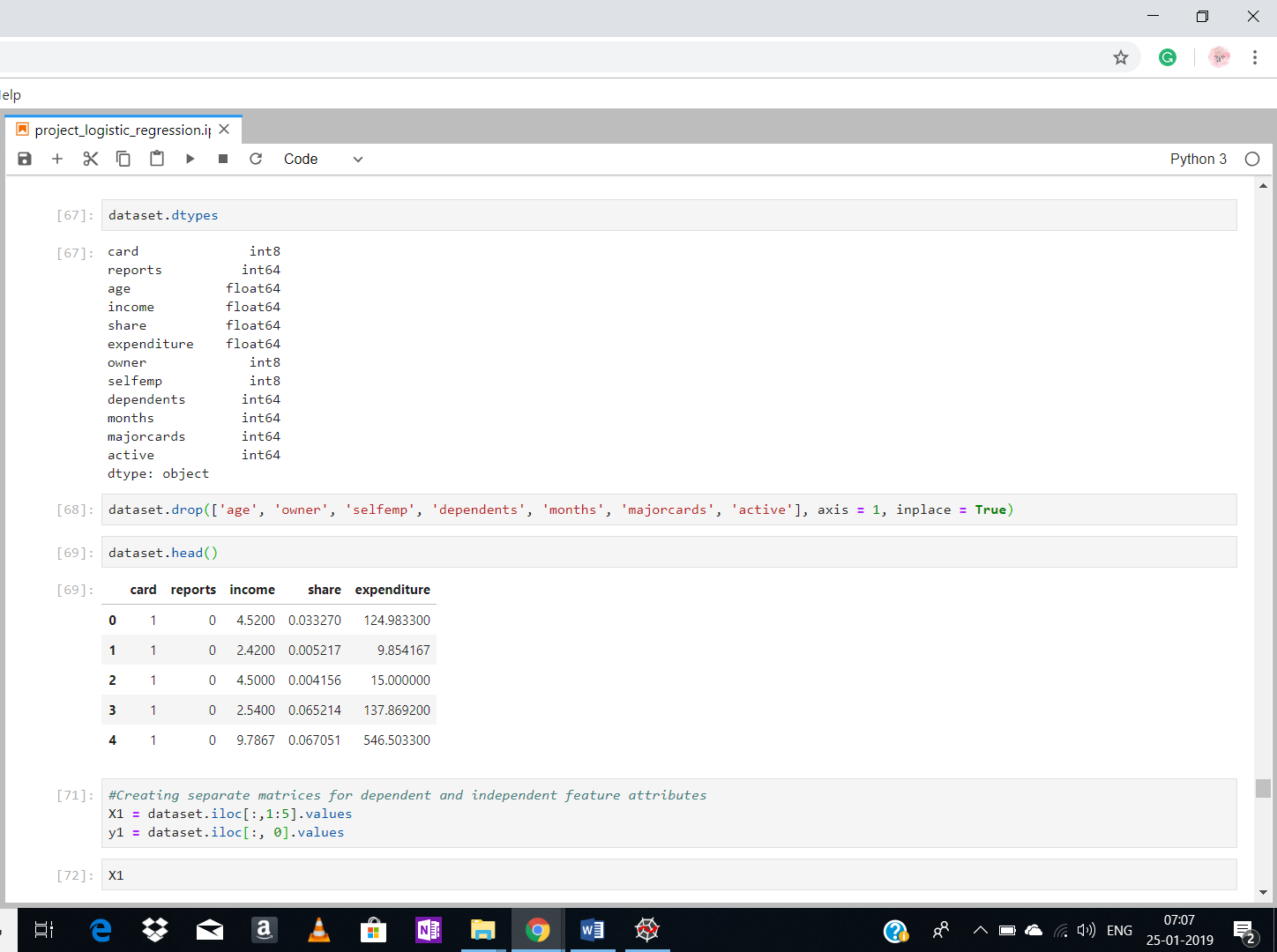


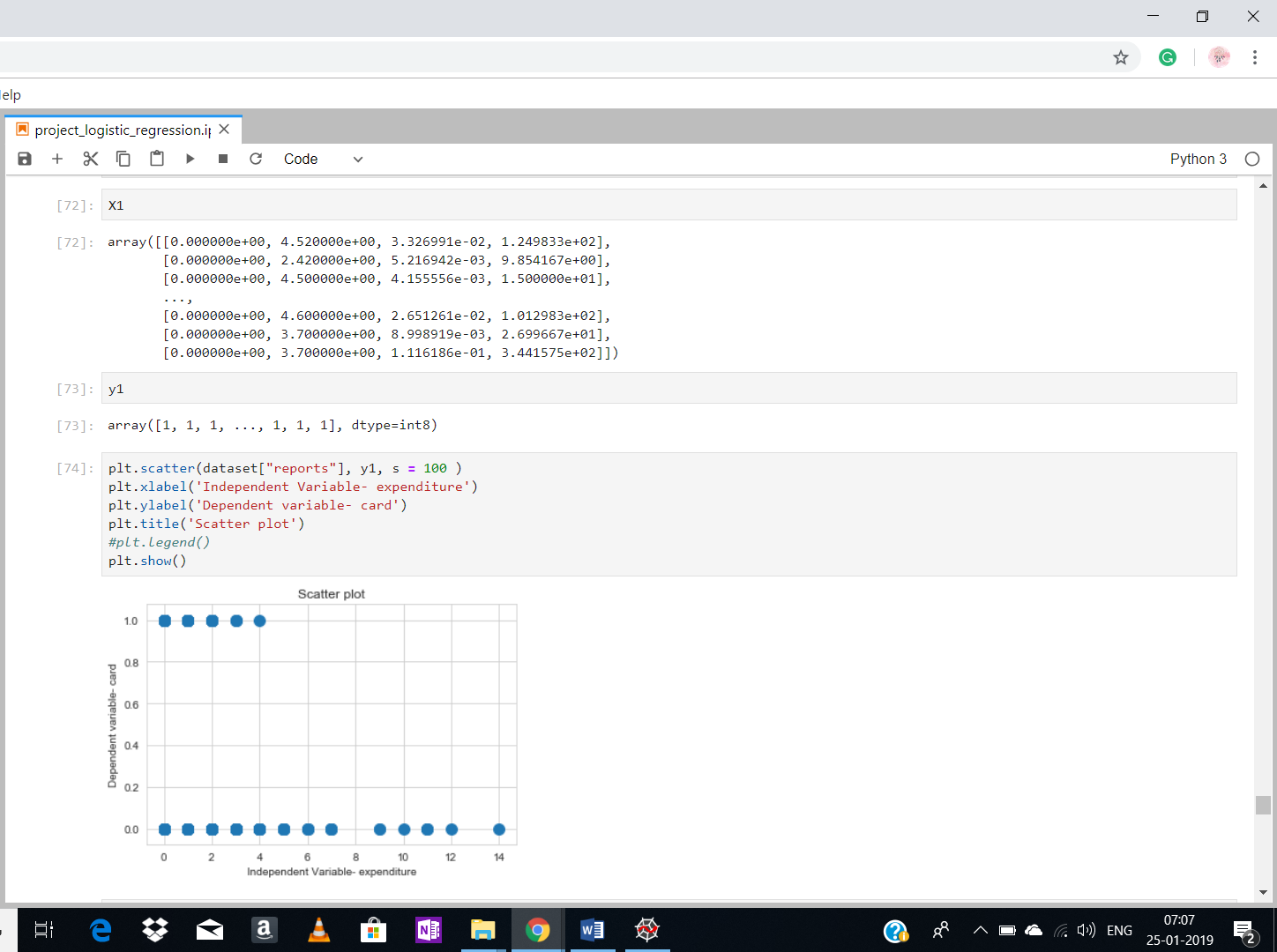


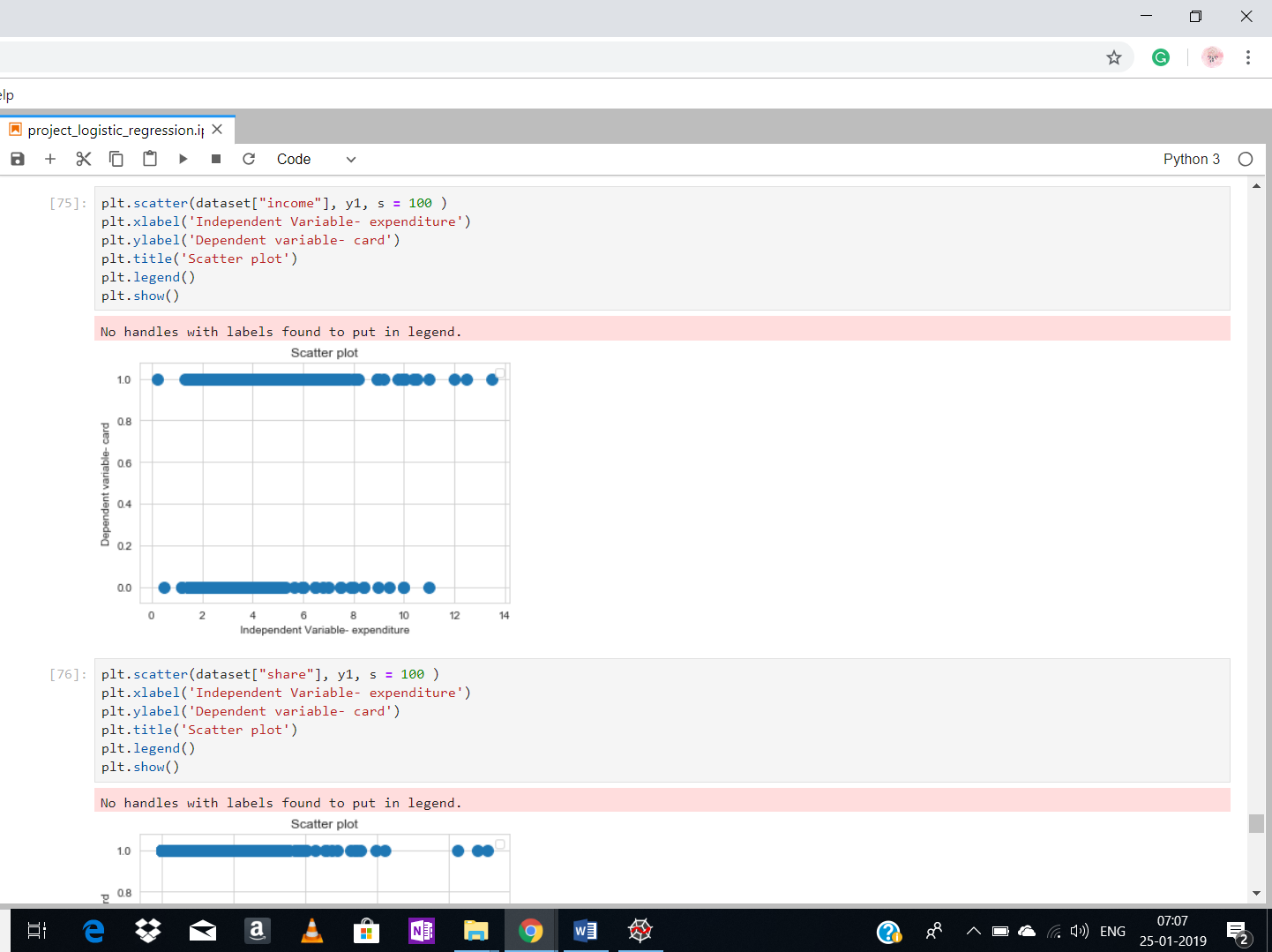


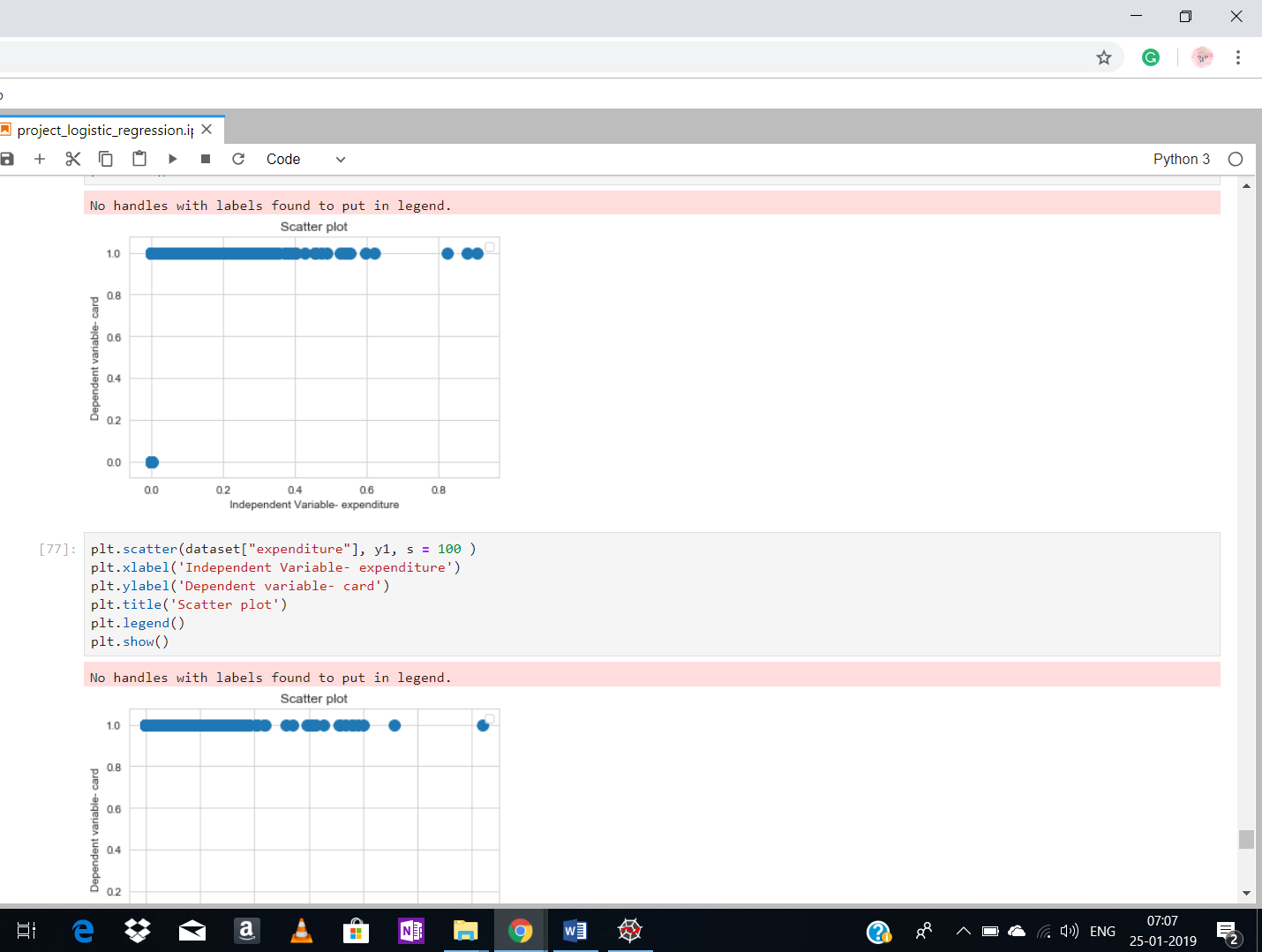


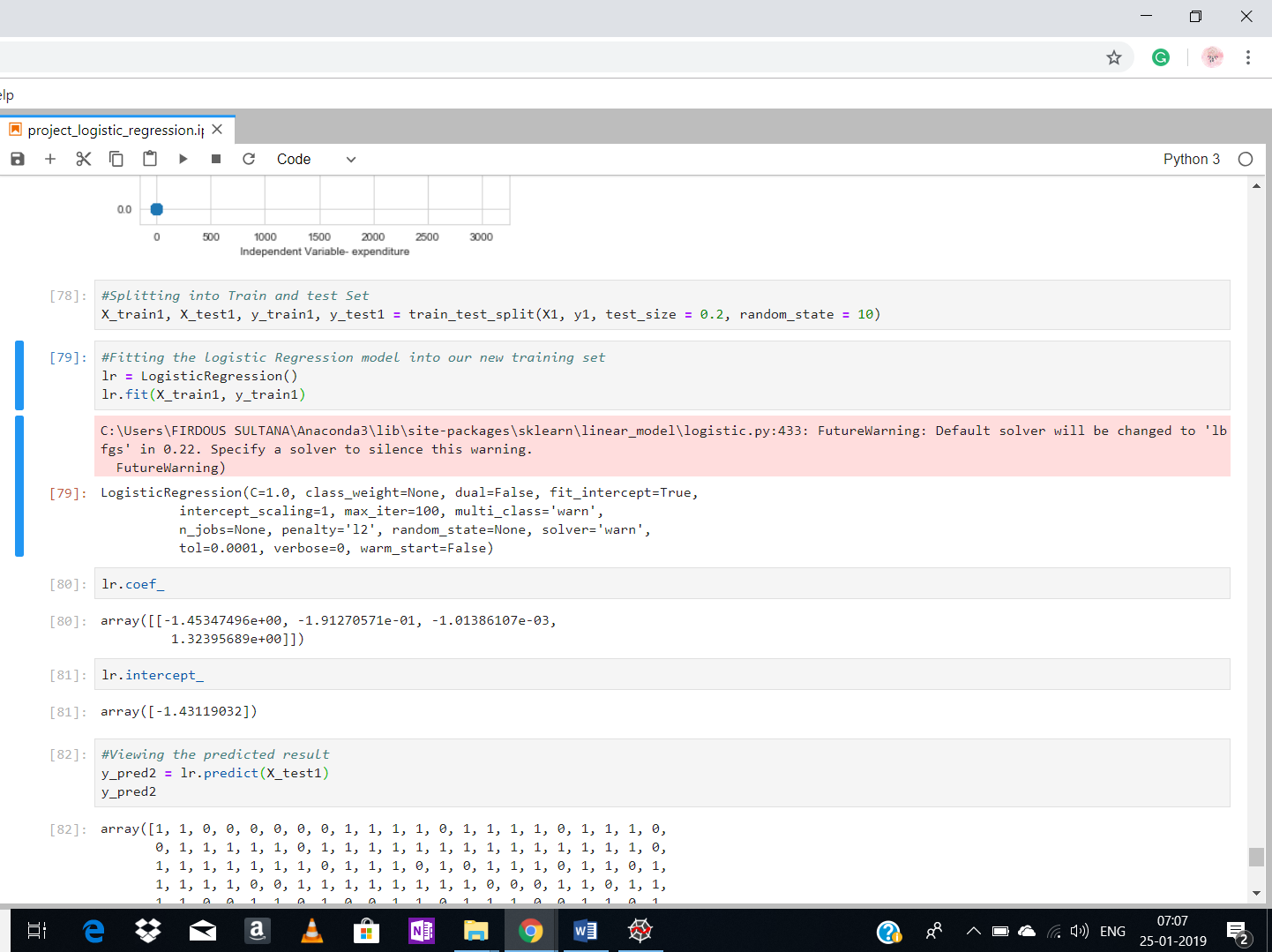


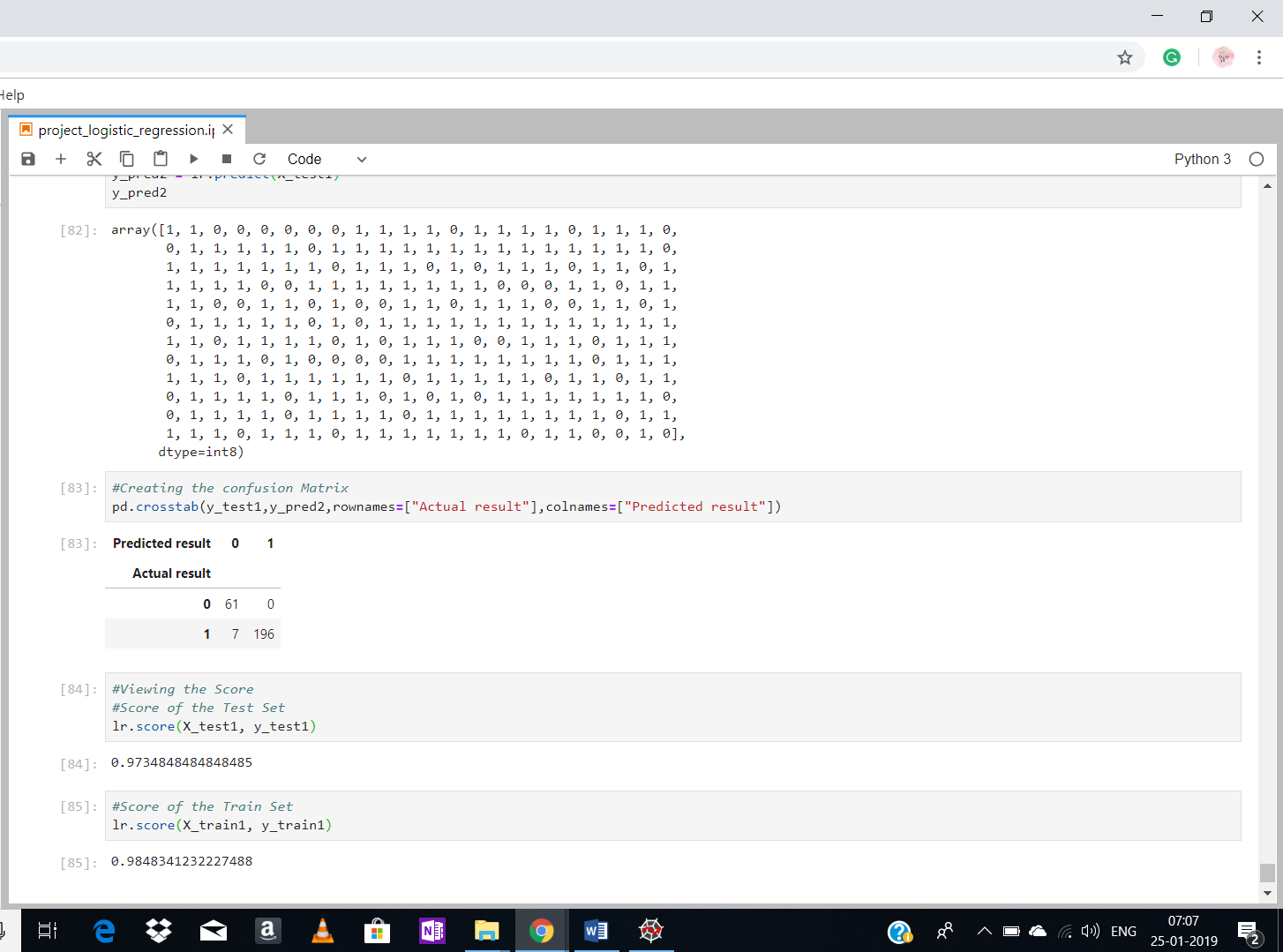












Comparison chart between the models:



Order of Accuracy:

1. Decision Tree 🡪 97.97%
2. Naïve Bayes 🡪 97.72%
3. Logistic Regression 🡪 97.34%
4. K- NN 🡪 96.96%

future scope of improvements:

* In the dataset, every “No” value of card in the dataset has “0” expenditure, which causes ambiguity. It would be better to predict well if expenditures are not like this.
* The decision Tree is more simple to visualize, but I came across an ambiguity that how much depth is to be kept so that we always arrive at some result in each case. This requires various trial methods.

## bibliography:

* <https://youtu.be/qDcl-FRnwSU>
* <https://machinelearningmastery.com/k-nearest-neighbors-for-machine-learning/>
* <https://en.wikipedia.org/wiki/Naive_Bayes_classifier>
* <https://en.wikipedia.org/wiki/Logistic_regression>